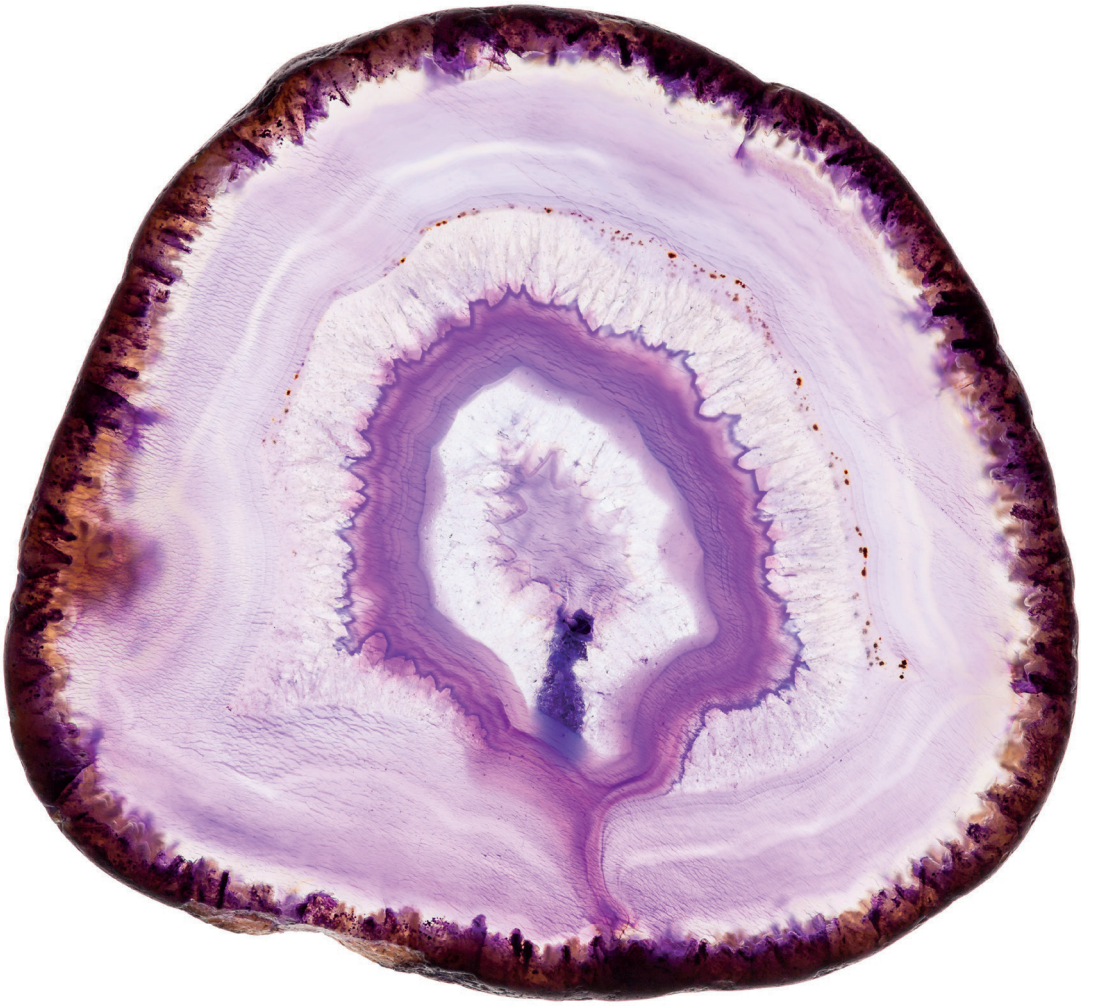
A multi-omics view of Pancreatic Cancer

Mantini, Giulia

2021

document version
Publisher's PDF, also known as Version of record

A multi-omics view of Pancreatic Cancer



Giulia Mantini

Giulia Mantini

A multi-omics view of Pancreatic Cancer

ISBN: 978-94-6423-325-4
Cover design & lay-out: Lisa Plets (proefschriftmaken) & Giulia Mantini
Printed by: ProefschriftMaken || www.proefschriftmaken.nl

The cover image represents a cross-section of a mineral, the agate.

This research project was funded by KWF Kanker berstrijding, AIRC foundation (Associazione Italiana per la Ricerca sul Cancro), Stichting VUmc CCA and Fondazione Pisana Per La Scienza.

VRIJE UNIVERSITEIT

A multi-omics view of Pancreatic Cancer

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor of Philosophy
aan de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. V. Subramaniam,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Geneeskunde,
op dinsdag 6 juli 2021 om 11.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door
Giulia Mantini
geboren te Pescara, Italië

promotoren:	prof.dr. C. R. Jimenez prof.dr. T. Würdinger
copromotoren:	dr. E. Giovannetti dr. M.F. Bijlsma
promotiecommissie:	prof.dr. R.D.M. Steenbergen prof.dr. F. Beltram prof.dr. J. Heringa dr. T.V. Pham dr. A.E. Frampton dr. C. Bachas

CONTENTS

Chapter 1

General introduction and thesis outline

Epidemiology and pathology of pancreatic cancer

Recent epidemiological studies reported an increasing trend in the incidence and mortality rates for pancreatic cancer. Based on global estimations (GLOBOCAN 2018), pancreatic cancer has caused 4.5% of all death caused by cancer in 2018[1,2,3].

The poor prognosis of pancreatic cancer is attributable to non-specific symptoms leading to diagnosis at a late stage, the absence of effective biomarkers for early detection of disease, aggressive metastatic behavior, early recurrence, lack of effective treatments, and development of drug resistance.

Although the major causes of pancreatic cancer are still unclear, certain risk factors such as obesity, diabetes mellitus, age, family history, and chronic pancreatitis have been identified[4].

The most prevalent type of pancreatic cancer is pancreatic ductal adenocarcinoma (PDAC), which accounts for more than 90% of all pancreatic malignancies[5]. Most often, PDAC originates through a process called "acinar-to-ductal metaplasia", where pancreatic acinar cells differentiate into ductal-like cells with ductal cell traits. This process favors pancreatic acinar cells to adapt to oncogenic changes and environmental stress, often leading to pancreatic intraepithelial neoplasia (PanIN), a common precancerous lesion that precedes pancreatic cancer. PDAC commonly arises in the head of the pancreas, whereas the involvement of the pancreatic body or tail is rarer[6]. Morphologically, PDAC presents as a solid poorly-defined mass, typically between 2–4 cm at diagnosis, which infiltrates surrounding structures (peripancreatic adipose tissue, duodenum, stomach, portal vein, and regional lymph nodes). Microscopically, PDAC consists of atypical tubular glands but can include non-tubular components, and growth patterns are strikingly heterogeneous between and within tumors[7]. Moreover, PDAC epithelial cancer cells are embedded in a prominent desmoplastic reaction, known as the "stroma". The stroma is mainly constituted by cancer-associated fibroblasts (CAFs), immune cells (T & B cells, NK cells, tumor-associated macrophages), blood vessels, extracellular matrix, and a liquid milieu of cytokines, growth factors, and exosomes[8,9]. Together, epithelial and stromal cells form the tumor microenvironment (TME). An orchestrated cross-talk between epithelial and stromal cells in the TME can stimulate proliferation and metastatic growth alongside drug resistance and tumor relapse.

Regardless of the multitude of studies recognizing epithelial-stromal interactions as important targets for cancer therapy, identifying relevant and 'druggable' cancer- stromal interactions remains challenging due to the lack of quantitative methods to analyze the whole cancer-stromal interactome.

Extensive efforts have been made to investigate novel diagnostic, prognostic, predictive markers, and drug targets that could optimize and guide novel treatment regimens for PDAC. However, this disease remains lethal and exhibits the lowest survival rate among all cancers, where only 9% of patients survive beyond 5 years[10].

Current diagnostic procedures and therapeutic treatments

At present, early diagnosis at a curable stage is challenging. Patients rarely exhibit early symptoms, and most of these are non-specific (jaundice, sudden weight loss, fatigue).

Moreover, this type of tumor does not have sensitive and specific markers to aid detection. There is only a paucity of accurate molecular tools for diagnosis, prognosis, and treatment selection. Carbohydrate antigen 19-9 (CA 19-9), an antigen measured in blood, is the only diagnostic biomarker for pancreatic cancer approved by FDA and can be used in conjunction with CT scan and MRI. Measurement of blood CA 19-9 levels can also be used to evaluate treatment response after surgery or chemotherapy[11]. However, CA 19-9 is not recommended in a screening setting due to inadequate sensitivity and specificity. Indeed, increased levels of CA 19-9 can be caused by other pancreaticobiliary diseases, such as chronic pancreatitis or biliary obstruction, resulting in a false-positive response. Thus, more sensitive and specific diagnostic markers for PDAC are urgently needed, such as non-invasive biomarkers that can be detectable in accessible biofluids such as blood, urine, and saliva.

Liquid biopsies consist of sampling and analysis of biofluids that allow doctors to collect samples for diagnosis and molecular follow-up in a non-invasive manner. Liquid biopsies are already in every diagnostic routine and have also been implemented in the clinic to guide and monitor the therapies of several cancer patients[12,13]. Multiple recent studies support the ability of blood liquid biopsies to distinguish PDAC from healthy individuals[14,15], and emerging pieces of evidence are reporting the role of blood platelets and PDAC.

PDAC is traditionally associated with thrombocytosis and hypercoagulation[16–18]. More than 150 years ago, the physician Armand Trousseau described spontaneous coagulation in cancerous patients for the first time. After suffering from several migratory venous thromboses, he diagnosed himself with gastric cancer and died the year later of pancreatic cancer[19]. Yet, a proven relationship between PDAC and blood platelets is missing, and only recently, blood platelets are receiving more attention as an alternative source for cancer biomarkers.

Platelets have a lifespan of 7-10 days in the circulation and may act as responders to the presence of cancer[20]: sequestering proteins or RNAs from tumor apoptotic cells that can educate the platelet itself; altering its RNA profile via splicing[21,22]. The resulting platelet is then called tumor educated platelet (TEP)[23]. So far, it has been shown that TEPs from patients with cancer are distinct from those with inflammatory and other noncancerous diseases and can be potentially applied in the clinic for cancer diagnostic[22,24].

Early diagnosis of PDAC is a crucial step to improve the efficacy of treatment and patient outcomes. Approximately only 20% of patients are diagnosed in early stages and are amenable to curative surgery, while 80% have locally advanced, non-resectable tumors or tumors with distant metastases[25,26]. Therefore, the vast majority of patients are treated upfront with systemic therapy. Current guidelines mention adopting systemic chemotherapy with multi-drug regimens. For a long time, the treatment was limited to gemcitabine with minimal benefit but improved quality of life (QoL). In 2011 and 2013, results of two phase III studies showed a statistically relevant and clinically meaningful improvement in the overall survival of metastatic PDAC patients, thus providing two first-line chemotherapeutic options: (i) 5-fluorouracil/leucovorin combined with irinotecan and oxaliplatin (FOLFIRINOX); (ii) the combination of gemcitabine and nab-paclitaxel compared with gemcitabine alone.

However, these protocols are associated with high toxicity and modest response rates[27].

Thus, significant efforts should be taken to identify both new treatments and new biomarkers to select subgroups of patients who may benefit from specific therapies.
To this end, the analysis of multiple molecular profiles has emerged as an unbiased strategy for the identification of novel drug targets and predictive biomarkers.
Below we will discuss the -omic profiles of PDAC analyzed in this thesis (**Figure 1**).

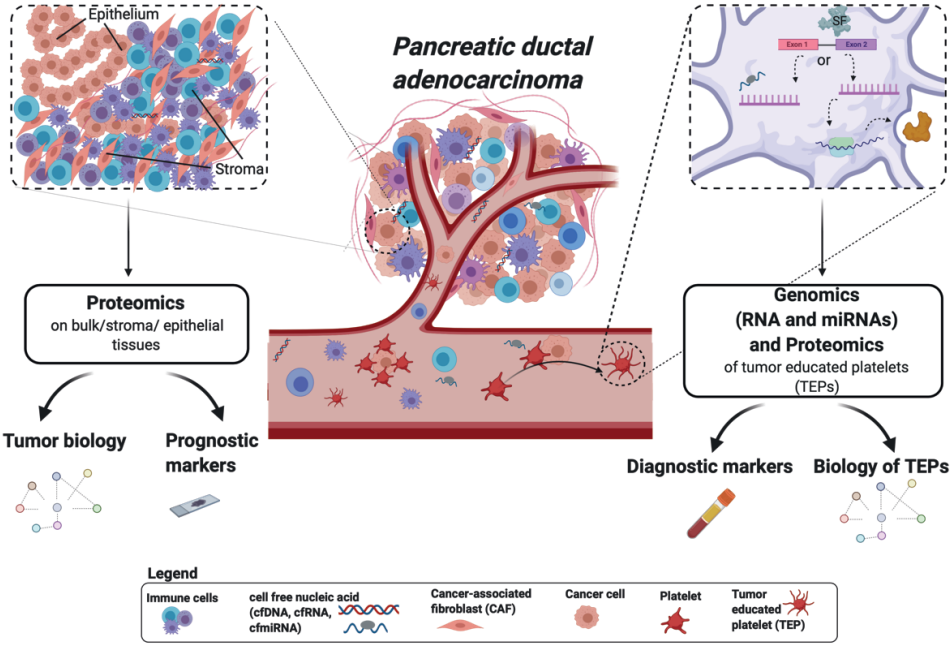


Figure 1. Schematic overview of the research studies described in this thesis.
Proteomics analysis from tissue biopsies is used to define new and additional prognostic markers as well as markers for therapy decisions. Transcriptomics analysis of tumor-educated platelets (TEPs) can be used for cancer diagnostic. Omics analysis of TEPs reveals biological mechanisms underlying tumor-platelet education.

OMICS APPROACHES IN PANCREATIC CANCER

In the wake of the Human Genome Project results, an increasing number of diagnoses were expected, including the boost for treatment and prevention in oncological diseases. The progress in genomic research has dramatically increased our knowledge on the molecular basis of most tumors, including PDAC. Several large-scale genomic studies showed PDAC stratification based on gene expression [28–30]. However, these findings have not yet been translated into clinical practice making the analysis of one -omic layer insufficient to get insights on cancer phenotype. Further investigation and integration of different data types (e.g., genomics, proteomics, epigenomics, and metabolomics) can contribute to the identification of specific PDAC subgroups and improve the clinical management of PDAC patients.

Mutational status

PDAC treatment has remained elusive partly due to the absence of targetable mutations. The genetic of PDAC consists of a pattern of high penetrance genetic alterations that occur in four genetic loci: KRAS, TP53, CDKN2A, and SMAD4/DPC4 [29–33]. The most frequent alteration is the proto-oncogene KRAS (over 95% of cases, far exceeding the rate of any other cancer). This aberration is likely to be the main driving force behind PDAC disease, having been identified in very early, low-grade neoplastic precursor lesions. This alteration was considered undruggable in the past, but newer therapies that directly block KRAS specific mutations are now in clinical trials. Besides, activation of KRAS oncogene is accompanied by inactivation of tumor suppressors such as p16, TP53, and SMAD4. However, many other genetic changes occur at lower frequencies, such as mutations in the splicing factor 3B subunit 1 (SF3B1)[29]. A recent study suggests a link between high levels of SF3B1 and cancer aggressiveness, although the underlying mechanisms for this phenomenon remain elusive[34]. These genomic changes are suggestive of significant genomic instability[34,35] and may also limit the effectiveness of therapy by contributing to secondary/acquired chemoresistance.

MicroRNA profiling

Emerging shreds of evidence suggest that microRNA (miRNA) expression patterns can have an important role as biomarkers in pancreatic cancer[36,37]. MiRNAs are non-coding RNAs of 19-24 nucleotides. Most miRNA genes are transcribed in the nucleus, cleaved by Drosha protein, and exported in the cytoplasm, where they are further cleaved by the protein Dicer. Commonly, one strand of the miRNA is incorporated in the RNA-induced silencing complex (RISC) that will repress gene transcription at the mRNA target level[38]. In contrast, the other strand is often degraded. Importantly, malfunctions on Drosha and Dicer proteins or the addition of nucleotides to the 3' or 5' ends by nucleotidyltransferases can develop a modified form of miRNAs. Those newly discovered types of miRNAs (isomiRs) have already been described as potential biomarkers for prostate cancer detection[39] and cell type[40].

A large body of evidence has revealed that miRNAs are aberrantly expressed in the serum and cancer tissues. They are relatively easily detected by next-generation sequencing (NGS) technologies. Of note, many of the pathways in which miRNAs contribute to carcinogenesis can also be targets of anti-cancer drugs. In particular, many studies revealed that PDAC presents twelve altered core signaling pathways in approximately 80% of the cases[41,42]. The involvement of miRNAs as regulators of these pathways in various processes makes them an attractive field of research for innovative cancer diagnostics and precision medicine approaches.[41,43]

Transcriptomics and RNA splicing

Several studies have evaluated the transcriptional profiles of PDAC to stratify patients for personalized treatment strategies[30,32,33,44,45]. These studies showed that PDAC is a highly heterogenic disease with different subtypes and survival rates. Available transcriptome-based classifications differ in the numbers of subtypes identified. Nevertheless, all these studies share

common subtypes, including a classical/canonical subtype hallmarked by epithelial-like gene expression and a quasi-mesenchymal/basal-like subtype characterized by a more mesenchymal gene expression pattern and poorer prognosis. Notably, some of these stratification studies analyzed both tumoral and stromal compartments. These studies provided important clues on the tumor microenvironment of PDAC, which appeared to be involved in cancer development, cancer progression, and drug resistance.
However, there is still no consensus classification to date, which would be the prerequisite for clinical application in precision medicine[46,47].
A major contributor to transcriptome diversification is the splicing of RNA.
RNA splicing is an essential process in which protein complexes named as splicing factors (SFs) remove introns (non-coding regions) from the pre-mRNA sequence, joining the residual exons (coding region) together. The resulting RNA sequence is known as mRNA (mature messenger RNA) and is then translated into protein.
In many cases, the splicing process can create a range of unique proteins (protein isoforms) by varying one mRNA molecule's intron/exon composition. This alternative mechanism is called alternative splicing (AS).
Alternative splicing is one of the key regulatory events leading to transcriptome and proteome diversity, and so, different cancer phenotypes. The functional consequences of most AS events are still unknown. The encoded protein can result in a non-functional product or some splice variant product(s) that assume a completely different function from the wild-type protein. In some cases, the protein even acquires antagonistic functions[48].
A recent study on PDAC tissues showed that microarray analyses on splice variants were able to detect splice aberrations event[49]. Specifically, in this study, the emerging aberrations predominantly affect the extracellular matrix-associated genes and focal adhesion genes, which are commonly related to stromal compartment and poor survival. Therefore, more investigations on AS events are warranted, as they can play an important role as diagnostic and therapeutic targets.
However, retention introns, a rare event in normal eukaryotic tissues, seemed to be also hijacked in PDAC evolution[49]. Inclusion of intronic sequences within the mRNA (sometimes referred to as the phenomenon of exonization) has also been reported in cancer[50] and extensively studied in tumor-educated platelets[51,52].

Proteomics

The proteome is a vital complement to genomic data as genetic changes alone might be insufficient to understand complex and heterogeneous tumors. It provides valuable information about the downstream consequences of an aberrant genome and yields direct insight into cellular signaling states and the driver molecules of disease. A recent comparison of transcriptomics and proteomics data from colorectal, breast, and ovarian cancer data revealed that proteomics outperformed transcriptome profiling when predicting for gene function[53].

In PDAC proteomics, a wide variety of clinical specimens have been mainly analyzed in small-scale studies, including tissue samples, blood, pancreatic juice, cyst fluid, urine and bile fluid, cell lines, and patient-derived xenograft (PDX) models. Remarkably, few genomic studies have tried to investigate the TME cross-talk using laser capture microdissection technique (LCM),[30,54] indicating partial or null dependence between epithelial and stromal compartment. However, these analyses have been challenging due to differences in used platforms, inclusion criteria, and sample preparations resulting in a lack of consistency within the study. Most recently, with the advances in proteomics, LCM can be coupled to mass spectrometry[55,56]. This approach is particularly attractive in discovering new prognostic and predictive biomarkers using proteomics data and may boost the discovery of mechanisms underlying tumor-stroma cross-talks.

Recent advances in high-throughput mass spectrometry have improved quality and quantity of proteomics data facilitating cancer research. In addition, several computational methods already implemented in genomics can now be applied to proteomics data. An example is the application of weighted gene co-expression analysis (WGCNA) to proteomics data[57–59]. Correlation network analysis is emerging as a powerful approach to infer tumor biology from a set of single or multiple expression profiles. Weighted gene co-expression network analysis (WGCNA) assumes that the strength of node-to-node connections is best quantified by measures derived from their correlations. Constructing co-expression networks is an effective way to characterize correlation patterns among nodes (genes or proteins) and infer new biological functions of densely interconnected nodes called "modules". Modules can be related to external sample traits such as patient survival, recurrence, and disease/health state, to discover biomarkers or therapeutic targets.

Bioinformatics advances in data integration and drug combination

Recently, multiple studies showed that proteogenomic integration prioritized known genomics targets and yielded novel findings for therapeutic hypothesis generation[60–65]. A recent study of early-onset gastric cancer from Mun and colleagues[64] demonstrated successful integration of transcriptomics, proteomics, phosphoproteomics, and N-glycoproteomics profiles for cancer subtype stratification. The study indicates that two subtypes were mainly identified by proteomics and phosphoproteomics data and were associated with best and worst survival. Furthermore, drug sensitivity was predicted based on phosphoproteome information.

Considering all these features, data integration ensures to maximize biological insight for diagnostic, prognostic, and predictive markers, moving toward precision medicine.

Relevant to drug combination is the use of machine learning algorithms. For this purpose, several machine learning methods are now being applied to assess drug combination therapy for HIV, hypertension, infectious diseases, and cancer[66]. Current predictive models for drug-combination effects are commonly based on high-throughput testing of drug combinations for each cell line. These data are then used to identify the molecular features that predict therapy response[67]. A pan-cancer DREAM community challenge, the AstraZeneca Drug Combination Prediction Challenge, showed such findings on a pan-cancer scale[68]. Synergy prediction based

on models of drug interactions is appearing. These computational models should provide sufficient complexity to predict the effect of combination therapies on a cellular/signaling level accurately. Given this high complexity, more advanced machine-learning methods, such as deep learning, might be needed to enable adequate modeling.

AIM AND OUTLINE OF THE THESIS

The analysis of multiple -omics data is increasingly attracting the attention of the research community. It could provide important information and cover gaps that are often hidden or not seen when analyzing one single data type at a time.

This thesis aims to the underlying mechanisms of PDAC, analyzing different types of molecules at single or multiple -omics layers with a particular focus on proteome and miRNome. The overall goal is to provide improved insights into PDAC biology and novel candidate markers for diagnosis, prognosis, and treatments.

PART ONE – DISCOVERY OF NOVEL BIOMARKERS FROM PROTEOMICS

The first part of this thesis describes several protein markers to accurately stratify PDAC patients with good and poor prognoses after surgical resection. Those markers were analyzed from stromal and epithelial micro-dissected tissues and bulk tissues, and successfully validated on tissue microarrays (TMA) by immunohistochemistry and immunofluorescence. Additionally, this part provides an extensive description of the state of the art of integrative analysis using phosphoproteomics.

In particular, in **Chapter 2,** we reviewed integrative tools, algorithms, and existing pipelines, with a special focus on integrative methods that include phosphoproteomics data. Additionally, already existing methods are described and presented as future applications for phosphoproteomics data integration.

Chapter 3 explored the use of co-expression networks to provide insights into PDAC biology and discover new protein markers for PDAC prognosis after surgical resection. The knowledge of prognostic markers was expanded in **Chapter 4,** where LCM was coupled to LC-MS/MS to enrich stromal and epithelial compartments. The resulting prognostic markers for each compartment were evaluated and validated. Moreover, the proteomics profile of stromal compartment was recapitulating previous stromal subtypes evaluated by transcriptomics data.

PART TWO - INNOVATIVE ANALYSES OF MiRNAs AND RNAs

The second part of this thesis examines the important role of miRNAs and their modified forms (isomiRs), as well as intron retained RNA transcripts, to decipher the complex biology behind PDAC and exploit new potential miRNAs in PDAC diagnosis and treatment. To this end, in **Chapter 5**, we provided an overview of targetable miRNAs involved in aberrant PDAC metabolism.

Chapter 6 describes the use of a machine learning method for PDAC staging, which can be helpful for clinicians to understand the correct strategy to adopt at the time of diagnosis.

Chapter 7 explored the biology of tumor-educated platelets using multiple -omics layers (small RNAs, transcriptomics, and proteomics). This study confirmed the splicing activity on PDAC platelets and the enrichment of specific modified forms of miRNAs in PDAC tumors when compared to benign lesions. Moreover, this study provides a new catalog of multi-omics data for PDAC platelets.

PART THREE – COMPUTATIONAL TOOLS IN PHARMACOLOGICAL STUDIES

The majority of PDAC patients are diagnosed at a late stage or when the tumor has already distant metastases. These patients are treated upfront with chemotherapy, but most patients do not respond to the treatment and the development of acquired resistance is common. This emphasizes the importance of predictive biomarkers for treatment response and the need for new computational methods to define better drug combination strategies.

Actual strategies and future applications for drug combination are extensively discussed in **Chapter 8.** While **Chapter 9** describes the use of a transmembrane protein as a predictive biomarker for gemcitabine uptake and the subsequent important application in personalized medicine.

Splicing inhibitors are only recently receiving attention for PDAC treatments. **Chapter 10** reviewed several splicing factors aberrantly expressed in PDAC and their relationships with their target miRNAs. **Chapter 11** present a study focusing on one of the most mutated splicing factor genes, SF3B1, in PDAC and its inhibitors, evaluating PDAC tissue samples and preclinical models.

Chapter 12 provides a general discussion, summarizing the insights obtained from Chapter 2 to 11 and future perspectives.

References

1. Quante, A. S. *et al.* Projections of cancer incidence and cancer-related deaths in Germany by 2020 and 2030. *Cancer Med.* **5**, 2649–2656 (2016).
2. Rawla, P., Sunkara, T. & Gaduputi, V. Epidemiology of Pancreatic Cancer: Global Trends, Etiology and Risk Factors. *World J. Oncol.* **10**, 10–27 (2019).
3. Rahib, L. *et al.* Projecting Cancer Incidence and Deaths to 2030: The Unexpected Burden of Thyroid, Liver, and Pancreas Cancers in the United States. *Cancer Res.* **74**, 2913–2921 (2014).
4. Pancreatic Cancer Risk Factors. https://www.cancer.org/cancer/pancreatic-cancer/causes-risks-prevention/risk-factors.html.
5. Kleeff, J. *et al.* Pancreatic cancer. *Nat. Rev. Dis. Primer* **2**, (2016).
6. Haeberle, L. & Esposito, I. Pathology of pancreatic cancer. *Transl. Gastroenterol. Hepatol.* **4**, 50–50 (2019).
7. Cros, J., Raffenne, J., Couvelard, A. & Poté, N. Tumor Heterogeneity in Pancreatic Adenocarcinoma. *Pathobiology* **85**, 64–71 (2018).
8. Balkwill, F. R., Capasso, M. & Hagemann, T. The tumor microenvironment at a glance. *J. Cell Sci.* **125**, 5591–5596 (2012).
9. Quail, D. F. & Joyce, J. A. Microenvironmental regulation of tumor progression and metastasis. *Nat. Med.* **19**, 1423–1437 (2013).
10. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2020. *CA. Cancer J. Clin.* **70**, 7–30 (2020).
11. Duffy, M. J. *et al.* Tumor markers in pancreatic cancer: a European Group on Tumor Markers (EGTM) status report. *Ann. Oncol.* **21**, 441–447 (2010).
12. Chudasama, D. *et al.* Liquid Biopsies in Lung Cancer: Four Emerging Technologies and Potential Clinical Applications. *Cancers* **11**, 331 (2019).
13. Stefanovic, S. *et al.* Molecular Subtype Conversion between Primary and Metastatic Breast Cancer Corresponding to the Dynamics of Apoptotic and Intact Circulating Tumor Cells. *Cancers* **11**, 342 (2019).
14. Cohen, J. D. *et al.* Combined circulating tumor DNA and protein biomarker-based liquid biopsy for the earlier detection of pancreatic cancers. *Proc. Natl. Acad. Sci.* **114**, 10202–10207 (2017).
15. Cohen, J. D. *et al.* Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* **359**, 926 (2018).
16. Mandalà, M. *et al.* Venous thromboembolism predicts poor prognosis in irresectable pancreatic cancer patients. *Ann. Oncol.* **18**, 1660–1665 (2007).
17. Ishigaki, K. *et al.* Thromboembolisms in Advanced Pancreatic Cancer: A Retrospective Analysis of 475 Patients. *Pancreas* **46**, 1069–1075 (2017).
18. Ansari, D., Ansari, D., Andersson, R. & Andrén-Sandberg, Å. Pancreatic cancer and thromboembolic disease, 150 years after Trousseau. *Hepatobiliary Surg. Nutr.* **4**, 325–335 (2015).
19. Bariéty, M. [Tribute to Armand Trousseau (14 October 1801-23 June 1867)]. *Bull. Acad. Natl. Med.* **151**, 627–635 (1967).
20. McAllister, S. S. & Weinberg, R. A. The tumour-induced systemic environment as a critical regulator of cancer progression and metastasis. *Nat. Cell Biol.* **16**, 717–727 (2014).
21. Best, M. G., Vancura, A. & Wurdinger, T. Platelet RNA as a circulating biomarker trove for cancer diagnostics. *J. Thromb. Haemost.* **15**, 1295–1306 (2017).

22. Best, M. G. *et al.* RNA-Seq of Tumor-Educated Platelets Enables Blood-Based Pan-Cancer, Multiclass, and Molecular Pathway Cancer Diagnostics. *Cancer Cell* **28**, 666–676 (2015).
23. In 't Veld, S. G. J. G. & Wurdinger, T. Tumor-educated platelets. *Blood* **133**, 2359–2364 (2019).
24. Best, M. G., In 't Veld, S. G. J. G., Sol, N. & Wurdinger, T. RNA sequencing and swarm intelligence–enhanced classification algorithm development for blood-based disease diagnostics using spliced blood platelet RNA. *Nat. Protoc.* **14**, 1206–1234 (2019).
25. Gillen, S., Schuster, T., Meyer zum Büschenfelde, C., Friess, H. & Kleeff, J. Preoperative/Neoadjuvant Therapy in Pancreatic Cancer: A Systematic Review and Meta-analysis of Response and Resection Percentages. *PLoS Med.* **7**, e1000267 (2010).
26. Werner, J. *et al.* Advanced-stage pancreatic cancer: therapy options. *Nat. Rev. Clin. Oncol.* **10**, 323–333 (2013).
27. Gourgou-Bourgade, S. *et al.* Impact of FOLFIRINOX Compared With Gemcitabine on Quality of Life in Patients With Metastatic Pancreatic Cancer: Results From the PRODIGE 4/ACCORD 11 Randomized Trial. *J. Clin. Oncol.* **31**, 23–29 (2013).
28. Collisson, E. A. *et al.* Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. *Nat. Med.* **17**, 500 (2011).
29. Raphael, B. J. *et al.* Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. *Cancer Cell* **32**, 185-203.e13 (2017).
30. Moffitt, R. A. *et al.* Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nat. Genet.* **47**, 1168 (2015).
31. Witkiewicz, A. K. *et al.* Whole-exome sequencing of pancreatic cancer defines genetic diversity and therapeutic targets. *Nat. Commun.* **6**, (2015).
32. Waddell, N. *et al.* Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* **518**, 495–501 (2015).
33. Bailey, P. *et al.* Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* **531**, 47 (2016).
34. Biankin, A. V. *et al.* Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature* **491**, 399–405 (2012).
35. Notta, F. *et al.* A renewed model of pancreatic cancer evolution based on genomic rearrangement patterns. *Nature* **538**, 378–382 (2016).
36. Frampton, A. E. *et al.* microRNAs with prognostic significance in pancreatic ductal adenocarcinoma: A meta-analysis. *Eur. J. Cancer* **51**, 1389–1404 (2015).
37. Shams, R. *et al.* Identification of potential microRNA panels for pancreatic cancer diagnosis using microarray datasets and bioinformatics methods. *Sci. Rep.* **10**, (2020).
38. Bartel, D. P. MicroRNAs: Target Recognition and Regulatory Functions. *Cell* **136**, 215–233 (2009).
39. Koppers-Lalic, D. *et al.* Non‑invasive prostate cancer detection by measuring miRNA variants (isomiRs) in urine extracellular vesicles. *Oncotarget* **7**, (2016).
40. Koppers-Lalic, D. *et al.* Nontemplated Nucleotide Additions Distinguish the Small RNA Composition in Cells from Exosomes. *Cell Rep.* **8**, 1649–1658 (2014).
41. Jones, S. *et al.* Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* **321**, 1801–1806 (2008).
42. Stoica, A.-F., Chang, C.-H. & Pauklin, S. Molecular Therapeutics of Pancreatic Ductal Adenocarcinoma: Targeted Pathways and the Role of Cancer Stem Cells. *Trends Pharmacol. Sci.* **41**, 977–993 (2020).

43. Sun, L. *et al.* MicroRNA Signaling Pathway Network in Pancreatic Ductal Adenocarcinoma. *J. Genet. Genomics* **42**, 563–577 (2015).
44. Connor, A. A. *et al.* Association of Distinct Mutational Signatures With Correlates of Increased Immune Activity in Pancreatic Ductal Adenocarcinoma. *JAMA Oncol.* **3**, 774 (2017).
45. Dijk, F. *et al.* Unsupervised class discovery in pancreatic ductal adenocarcinoma reveals cell-intrinsic mesenchymal features and high concordance between existing classification systems. *Sci. Rep.* **10**, (2020).
46. Muckenhuber, A. *et al.* Pancreatic Ductal Adenocarcinoma Subtyping Using the Biomarkers Hepatocyte Nuclear Factor-1A and Cytokeratin-81 Correlates with Outcome and Treatment Response. *Clin. Cancer Res.* **24**, 351–359 (2018).
47. Aung, K. L. *et al.* Genomics-Driven Precision Medicine for Advanced Pancreatic Cancer: Early Results from the COMPASS Trial. *Clin. Cancer Res.* **24**, 1344–1354 (2018).
48. Sveen, A., Kilpinen, S., Ruusulehto, A., Lothe, R. A. & Skotheim, R. I. Aberrant RNA splicing in cancer; expression changes and driver mutations of splicing factor genes. *Oncogene* **35**, 2413–2427 (2016).
49. Wang, J. *et al.* Splice variants as novel targets in pancreatic ductal adenocarcinoma. *Sci. Rep.* **7**, (2017).
50. Braunschweig, U. *et al.* Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res.* **24**, 1774–1786 (2014).
51. Nassa, G. *et al.* Splicing of platelet resident pre-mRNAs upon activation by physiological stimuli results in functionally relevant proteome modifications. *Sci. Rep.* **8**, (2018).
52. Cimmino, G. *et al.* Activating stimuli induce platelet microRNA modulation and proteome reorganisation. *Thromb. Haemost.* **114**, 96–108 (2015).
53. Wang, J. *et al.* Proteome Profiling Outperforms Transcriptome Profiling for Coexpression Based Gene Function Prediction. *Mol. Cell. Proteomics* **16**, 121–134 (2017).
54. Maurer, C. *et al.* Experimental microdissection enables functional harmonisation of pancreatic cancer subtypes. *Gut* **68**, 1034–1043 (2019).
55. Shekouh, A. R. *et al.* Application of laser capture microdissection combined with two-dimensional electrophoresis for the discovery of differentially regulated proteins in pancreatic ductal adenocarcinoma. *PROTEOMICS* **3**, 1988–2001 (2003).
56. Le Large, T. Y. S. *et al.* Microdissected pancreatic cancer proteomes reveal tumor heterogeneity and therapeutic targets. *JCI Insight* **5**, (2020).
57. Seyfried, N. T. *et al.* A Multi-network Approach Identifies Protein-Specific Co-expression in Asymptomatic and Symptomatic Alzheimer's Disease. *Cell Syst.* **4**, 60-72.e4 (2017).
58. Zhang, Q. *et al.* Integrated proteomics and network analysis identifies protein hubs and network alterations in Alzheimer's disease. *Acta Neuropathol. Commun.* **6**, (2018).
59. Mantini, G. *et al.* Co-expression analysis of pancreatic cancer proteome reveals biology and prognostic biomarkers. *Cell. Oncol.* (2020) doi:10.1007/s13402-020-00548-y.
60. Drake, J. M. *et al.* Phosphoproteome Integration Reveals Patient-Specific Networks in Prostate Cancer. *Cell* **166**, 1041–1054 (2016).
61. Vasaikar, S. *et al.* Proteogenomic Analysis of Human Colon Cancer Reveals New Therapeutic Opportunities. *Cell* **177**, 1035-1049.e19 (2019).
62. Gao, Q. *et al.* Integrated Proteogenomic Characterization of HBV-Related Hepatocellular Carcinoma. *Cell* **179**, 561-577.e22 (2019).

63. Stewart, E. *et al.* Identification of Therapeutic Targets in Rhabdomyosarcoma through Integrated Genomic, Epigenomic, and Proteomic Analyses. *Cancer Cell* **34**, 411-426.e19 (2018).
64. Mun, D.-G. *et al.* Proteogenomic Characterization of Human Early-Onset Gastric Cancer. *Cancer Cell* **35**, 111-124.e10 (2019).
65. Komor, M. A. *et al.* Molecular characterization of colorectal adenomas reveals POFUT1 as a candidate driver of tumor progression. *Int. J. Cancer* **146**, 1979–1992 (2020).
66. Tsigelny, I. F. Artificial intelligence in drug combination therapy. *Brief. Bioinform.* **20**, 1434–1448 (2019).
67. NCI-DREAM Community *et al.* A community computational challenge to predict the activity of pairs of compounds. *Nat. Biotechnol.* **32**, 1213–1222 (2014).
68. Menden, M. P. *et al.* The germline genetic component of drug sensitivity in cancer cell lines. *Nat. Commun.* **9**, (2018).

PART ONE

DISCOVERY OF NOVEL BIOMARKERS FROM PROTEOMICS

Chapter 2

Computational Analysis of Phosphoproteomics Data in Multi-Omics Cancer Studies

Mantini G, Pham TV, Piersma SR, Jimenéz CR

Proteomics, 2020

REVIEW

Proteomics
Proteomics and Systems Biology
www.proteomics-journal.com

Computational Analysis of Phosphoproteomics Data in Multi-Omics Cancer Studies

*Giulia Mantini, Thang V. Pham, Sander R. Piersma, and Connie R. Jimenez**

Multiple types of molecular data for the same set of clinical samples are increasingly available and may be analyzed jointly in an integrative analysis to maximize comprehensive biological insight. This analysis is important as separate analyses of individual omics data types usually do not fully explain disease phenotypes. An increasing number of studies have now been focusing on multi-omics data integration, yet not many studies have included phosphoproteomics data, an important layer for understanding signaling pathways. Multi-omics integration methods with phosphoproteomics data are reviewed in the context of cancer research as well as multi-omics methods papers that would be promising to apply to phosphoproteomics data. Analysis of individual data types is still the major approach even in large cohort proteogenomics studies. Hence, a section is dedicated on possible integrative methods for multi-omics and phosphoproteomics data. In summary, this review provides the readers with both currently used integrative methods previously applied to phosphoproteomics and multi-omics data integration and other algorithms for multi-omics data integration promising for future application to phosphoproteomics data.

1. Introduction

In the era of comprehensive, unbiased molecular profiling, global analysis of post-translational modifications is of high interest for obtaining functional insights. One of the most common post-translational modifications that is involved in cell regulation and intracellular signal transduction is reversible protein phosphorylation catalyzed by protein kinases.[1] Through site-specific phosphorylation of protein substrates, kinases may modulate protein conformation, subcellular localization, protein–protein interaction, and signaling important for development, survival, and growth of all living beings. Importantly, signaling via tyrosine kinases is often deregulated in cancer as these enzymes mediate most growth and survival signaling in multicellular organisms.[2,3] This knowledge has led to the development of selective kinase inhibitors and important clinical achievements in cancer therapy. Examples of clinically successful kinase inhibitors include imatinib to block BCR-ABL fusion kinase in chronic myeloid leukemia,[4] trastuzumab for ERBB2 blockade in HER2-positive breast cancer,[5] crizotinib for ALK inhibition in anaplastic lymphoma,[6] and gefitinib to inhibit EGFR activity in lung cancer.[7]

In recent years, focus in precision medicine has shifted from identification of single genomic events to a more comprehensive assessment of tumor biology through integration of genomic, transcriptomic, and proteomic data.[8–10] In this context, integration of mass spectrometry-based phosphoproteome data is especially relevant, yet has only been explored in a limited number of studies.[11–17] Advances in the past decade in biochemical procedures for phosphopeptide enrichment, mass spectrometry technology, and computational approaches for phosphosite identification and quantification[18–20] have enabled the application of phosphoproteomics for the large-scale analysis of clinical materials. Currently, phosphoproteome data analysis and its integration into a multi-omics framework are challenging. Many tools simply collapse quantitative information from multiple phosphosites into a single protein phosphorylation value. This approach does not do justice to the site-specific phosphorylation behavior involved in signaling networks. Different sites in a protein have different functional consequences (e.g., activating vs inhibiting phosphosites) that need to be preserved in multi-omics data integration.

In this review, we describe current methods and tools for phosphoproteomics analysis as well as promising new methods for the integration of phosphoproteomics with other –omics data, taking into account the enormous complexity of the data.

2. Complexity of Phosphoproteome Data

Almost 500 000 phosphorylation sites are included in databases such as PhosphoSitePlus (PSP)[21] and approximately 500 kinases and 200 phosphatases are encoded in the human genome[22,23] that enable rapid regulation. Protein

G. Mantini, Dr. T. V. Pham, Dr. S. R. Piersma, Prof. C. R. Jimenez
Department of Medical Oncology
OncoProteomics Laboratory
CCA 1–60, Amsterdam UMC VUmc-location, De Boelelaan 1117,
Amsterdam 1081 HV, The Netherlands
E-mail: c.jimenez@vumc.nl

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/pmic.201900312

DOI: 10.1002/pmic.201900312

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

Proteomics
Proteomics and Systems Biology
www.proteomics-journal.com

phosphorylation leads to a reversible cascade of signaling events important for cell maintenance, proliferation, and survival. One kinase or phosphatase may have multiple substrates. These substrates can be phosphorylated at serine, threonine, or tyrosine residues.[24] Each site may be targeted by several kinases, depending on sequence context, resulting in an intricate network of kinases, target proteins, and phosphatases.[25] Bioinformatics tools such as PhosphoPredict,[26] NetworKIN,[27] and Phospho.ELM[28] are able to predict kinase-specific substrates based on motif analyses, yet functional annotation of the majority of phosphosites is incomplete. Thus, most research to date has been restricted to clustering and enrichment methods. These two methods group detected sites according to their quantitative behavior (clustering) and relate the mapped proteins to cellular processes like canonical pathways, transcriptional networks, or disease mechanisms via statistical testing against databases (enrichment). Another limitation is that modeling of phosphoproteomics networks is usually limited to few components (<50) where phosphosite information is collapsed into peptides or phosphoproteins[25] prior to obtaining functional annotation.[29] Thus, the phosphosite information is usually lost.

3. Current State-of-the-Art in Phosphoproteomics Data Analysis

3.1. Quantitative Phosphoproteomics

To enable optimal detection in large-scale mass spectrometry-based phosphoproteomics experiments, phosphopeptides need to be enriched from the cell lysate after tryptic digestion. Global phosphopeptide enrichment (phosphoserine, phosphothreonine, and phosphotyrosine) is achieved using affinity purification materials, including titanium dioxide beads (TiO_2)[30–32] or immobilized metal affinity chromatography (IMAC) resins like Ti-IMAC[33] and Fe(III)-IMAC.[34] Phosphotyrosine-containing peptides can be enriched specifically using selective sequence-context independent antibodies including 4G10, pY20, and pTyr-1000.[35] After enrichment, phosphopeptides are submitted to C18 nano-liquid chromatography coupled on-line to high-resolution/high mass accuracy tandem mass spectrometry for phosphopeptide identification and phosphosite localization. For phosphoproteomics, several fragmentation methods have been reported, including collision-induced (CID) fragmentation and higher-energy dissociation (HCD) that both suffer from neutral loss of threonine and serine phosphopeptides and electron-transfer dissociation (ETD) and combinations like EtHCD that result in more complete phosphopeptide sequence coverage.[36] Following LC-MS/MS data acquisition, peptides and phosphopeptides are identified by a search engine like MaxQuant[37] or Mascot.[38] Correct phosphosite localization, assigning a phosphorylation probability to each potential phosphosite (serine, threonine, or tyrosine) in a given phosphopeptide, requires optimized algorithms like A-score,[39] MaxQuant PTM-score,[40] PhosphoRS,[41] or Mascot delta score.[42] Upon phosphopeptide identification and phosphosite localization, differential analysis or pathway mapping requires phosphopeptide or site quantification. Several proteomics strategies are available for accurate and precise phosphorylation quantification, including label-free quantitation, stable-isotope in cell culture (SILAC), and isobaric labeling like iTRAQ and TMT, and their relative performance has been assessed showing label-free quantitation and SILAC to be most accurate in a mixed-species analysis while TMT MS2 resulted in the deepest data at the expense of accuracy due to ratio-compression.[20] Label-free quantitation is advantageous for large-scale clinical proteomics of heterogeneous samples and allows for single sample kinase activity analysis methods like integrative inferred kinase activity (INKA) discussed in Section 3.2.2 of this review. Recently, label-free phosphosite quantitation has been developed for data-independent analysis (DIA) and now includes a dedicated site localization algorithm that shows superior performance in phosphosite depth and quantitative accuracy over data-dependent analysis (DDA).[43]

3.2. Strategies for Kinase Activity Analysis

To determine kinase activities and nominate drug targets, several kinases ranking approaches have been previously described and can be characterized as group-based or single sample approach. For both approaches, dedicated data visualization is key.

3.2.1. Group-Based Analysis Tools

A substrate-centered approach to infer upstream kinase activity named kinase-enrichment analysis (KEA) was first reported by Lachmann et al.[44] in 2007 and further improved into kinase-substrate enrichment analysis (KSEA) to systematically infer the activation of given kinase pathways by Casado et al. in 2013. This computational approach was then converted into a web-based tool by Wiredja et al. in 2017.[45,46] To identify the substrates for each kinase, the KSEA App sources K-S annotations from PSP[21] by default, but users have the option to include NetworKIN annotations adjusting the inclusion stringency. In 2014, another tool named PSEA (phosphorylation set enrichment analysis) was released from Suo et al.,[47] where kinases characteristic of all collected disease-related phosphorylation substrates were analyzed using a kinase-specific prediction method. In the same year, Weidner et al.[48] developed PHOXTRACK (phosphosite-X-Tracing analysis of casual kinases). In contrast to the other tools, this user-friendly software uses the full set of quantitative proteomics data and applies non-parametric statistics to calculate differences in phosphosites based on different biological conditions. Recently, Mischnik and colleagues introduced a machine-learning approach named inference of kinase activities from phosphoproteomics (IKAP)[49] to estimate the activities of all kinases that are known to phosphorylate at least one phosphosite in a phosphoproteomics dataset. In contrast to KSEA, the output of IKAP is not only a score for the activity of a kinase, but also a value representing the strength of a kinase–substrate interaction in the investigated cell type.

Next to scoring kinase activity, visualizing kinases in a signaling pathway is also of great relevance. Pathway analysis of post-translational modifications (PTM) data is typically performed at a gene-centric level because of the lack of appropriately curated PTM signature databases. In 2019, Krug and colleagues presented the first version of PTM Signature Database (PTMSigDB).[50] The database contains a rich data source consisting of site-specific signatures of perturbations, kinase activities,

and signaling pathways curated from more than 2500 publications. The authors adapted the single sample Gene Set Enrichment Analysis approach to utilize PTMSigDB for a PTM Signature Enrichment Analysis (PTM-SEA) of quantitative MS data. PTMSigDB is the only available tool able to run a pathway enrichment analysis related to phosphorylated sites. The result is similar to a regular gene-set enrichment analysis where enriched pathways are ranked based on false discovery rate (FDR) score. The tool can only be applied for a group-comparison experiment.

3.2.2. Single-Sample Analysis Tools

To pinpoint hyperphosphorylated kinases as drug targets, Rikova et al.[2] performed large-scale phosphoproteomics of ≈200 lung cancer cell lines and clinical samples and sorted tyrosine kinases in single cancer samples on the basis of the sum of the spectral counts for all phosphopeptides attributed to a given kinase. The phosphokinases were then ranked by dividing the sum of phosphopeptide spectral counts by the number of samples with "non-zero" counts. Using this simple but effective approach, this pioneering 2007 study identified known and novel oncogenic kinases in lung cancer. This type of analysis can be performed in individual samples, but is limited by a focus on phosphorylation of the kinase itself, and does not include its substrates.

Recently, two tools have been reported to infer kinases activities in a single-sample manner: kinase activity ranking using phosphoproteomics data (KARP) (Wilkes et al., 2017) and INKA (Beekhof et al., 2019). KARP measures the net kinase activity as the sum of the intensities of its known substrates relative to the sum of the intensities of all phosphorylation sites present in the dataset. The result of such analysis is called *K-score*. INKA combines label-free kinase-centric (phosphokinase and activation loop data) and substrate-centric information from NetworKIN and PhosphoSitePlus to rank kinase activity (INKA score) in a single sample manner. INKA has been extensively benchmarked and applied, demonstrating that INKA scoring can i) identify oncogenes in cancer cell lines with known genomic aberrations, ii) identify differential kinase activity in models (wild-type vs mutant, +/− drug) and clinical samples (pre- and on-treatment tumor needle biopsies), iii) be correlated to drug sensitivity data, and iv) enable drug selection. The latter was shown for patient-derived tumor xenografts with functional testing in organoids by Beekhof et al. 2019[51] and for AML cell lines by Van Alphen et al. 2020.[52] The algorithm, easily accessible by inkascore.org website or in R language, requires MaxQuant output tables and enables a per-sample visualization of ranked kinase activities as bar graph and kinase–substrate network. Moreover, benchmarking INKA and KARP on a set of cancer cell lines with drug sensitivity data, INKA outperformed KARP in assigning a highest score to the known drivers.

Besides the extensive attention to kinase–substrate enrichment and kinase activity ranking, the functional role of many phosphosites remains to be elucidated. Recently, Ochoa et al.[53] applied a machine learning approach to public datasets and constructed a curated functional landscape of > 100 K phosphosites. Moreover, Ochoa condensed 59 features indicative of proteomic, structural, regulatory, or evolutionary relevance into a single functional score that can be used to identify regulatory phosphosites across different molecular mechanisms, processes, and diseases. This first study can be the basis of future functional studies in the phosphoproteomics field.

3.2.3. Tools for Phospho-Data Visualization

Cytoscape[54] is the most widely used software for network analysis and visualization with a well maintained dedicated app store[55] with more than 500 applications, among which the new published Cytoscape app "Omics Visualizer"[56] is particularly useful for visualizing phosphoproteomics. Researchers and developers around the world have contributed new functionalities and tailored solutions for a broad range of network-related projects.

Raaijmakers et al. developed a Cytoscape plugin named "PhosphoPath" in 2015.[57] The app was designed to visualize and analyze quantitative proteome and phosphoproteome datasets by importing publicly available data and allowing user to visualize specific phosphosites. However, this app can only be used for group-comparison experiments and data input needs to be ready upfront following the rigorous app-guidelines. In the following years, Narushima et al. developed another Cytoscape app "PTMapper" to integrate PPI data with publicly available kinase–substrate relations at the resolution of phosphorylated amino acid residues.[58] One year later, "PTMOracle" was released from Tay et al.[59] allowing the user to visualize and analyze their own PTM data.

Besides Cytoscape, aforementioned tools such as INKA and PTM-SEA provide both phospho-data analysis and visualization (see Section 3.2.2 for details). Additional tools for phospho-data visualization and databases sources are listed in **Table 1**.

However, none of the above-mentioned tools reflect sufficiently the complexity of the phosphoproteomics layer when analyzing multi-omics data. In many –omics integration studies, results from phosphoproteomics data are usually shown alone to avoid complexity in the figure. After accurately describing the data integration method, a small and centered signaling pathway is usually shown, highlighting the expression of phosphoproteins but more often losing the phosphosite information. A rather complete description and visualization of multi-omics data centered to the phosphoproteome level and more in detail to the phosphosite information is shown in **Figure 1**. The figure represents a hypothetical interaction of kinase, phosphatase, and substrates with multi-omics information (protein and gene abundances, phosphoproteomics signaling, mutations, and methylations) making use of kinase activity annotation from already existing tools such as INKA.

4. Phosphoproteomics Analysis in a Multi-Omics Context

Recently, multiple studies showed that proteogenomic integration not only prioritized known genomics targets but also yielded to novel findings for therapeutic hypothesis generation.[11–15] It stands to reason that the integration of phosphoproteomics, with global proteomics and genomics, may provide greater clues to these signaling events than single omics.

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

Proteomics
Proteomics and Systems Biology
www.proteomics-journal.com

Table 1. Repositories and visualization tools for phosphoproteomics data.

Repositories and visualization tools for phosphoproteomics data				
Name	Computational platform	Description	Availability	Reference
PRIDE	Web	Proteomics	https://www.ebi.ac.uk/pride/archive/	EMBL-EBI & Elixir
LinkedOmics	Web	TCGA and CPTAC web portal	http://www.linkedomics.org/login.php	[97]
OmicsDI	Web	Contains data sets from 11 repositories in a common data structure	https://www.omicsdi.org/	
cBioPortal	Web	Data repository, analysis and visualization	https://www.cbioportal.org	[96]
TCPA	Web	Data repository, analysis and visualization	https://tcpaportal.org	[62]
PeptideAtlas repository	Web	Proteomics	http://www.peptideatlas.org/repository/	Institute for System Biology
PathVisio	Local	Integrative visualization of pathways	https://github.com/PathVisio/pathvisio/releases	Maastricht University
G6G	Web	Collection of commercially available and free tools for omics analysis	https://www.g6g-softwaredirectory.com	Al Shapiro G6G Tech
qPhos	Web	Curated quantitative phosphoproteome datasets in human tissues and cell lines from published literatures	http://qphos.cancerbio.info/index.php	[112]
PhosphoSitePlus	Web	Multiple type of phosphosite information	https://www.phosphosite.org/homeAction	Cell Signalling Technology
dbPAF	Web	Integrative database of protein phosphorylation in animals and fungi	http://dbpaf.biocuckoo.org/index.php	[113]
PTMapper	Cytoscape	Integrate PPI data with publicly available kinase–substrate relations at the resolution of phosphorylated amino acid residues	https://github.com/y-narushima/PTMapper	[58]
iPTMnet	Web	Integrative bioinformatics approach combining text mining, data mining, and ontological representation to capture rich PTM information, including PTM enzyme–substrate-site relationships, PTM-specific protein–protein interactions (PPIs), and PTM conservation across species	https: //research.bioinformatics.udel.edu/iptmnet/	[102]
x2kweb	web	X2K Web infers upstream regulatory networks from signatures of differentially expressed genes. X2K Web produces inferred networks of transcription factors, proteins, and kinases predicted to regulate the expression of the inputted gene list.	https://amp.pharm.mssm.edu/X2K/	[114]
phosphoPath	Cytoscape	Visualization and analysis of quantitative proteome and phosphoproteome datasets. PhosphoPath imports publically available data from BioGrid, PhosphositePlus, and Wikipathways. PhosphoPath further enables PTM site-specific visualization of information from quantitative time series.	http://apps.cytoscape.org/apps/phosphopath	[57]
PoGo	Local	PoGo allows the visualization of protein sequences and the corresponding gene with mutation, CNV, PTM, and isoforms in one single environment (IGV)	http://www.sanger.ac.uk/science/tools/pogo	[115]

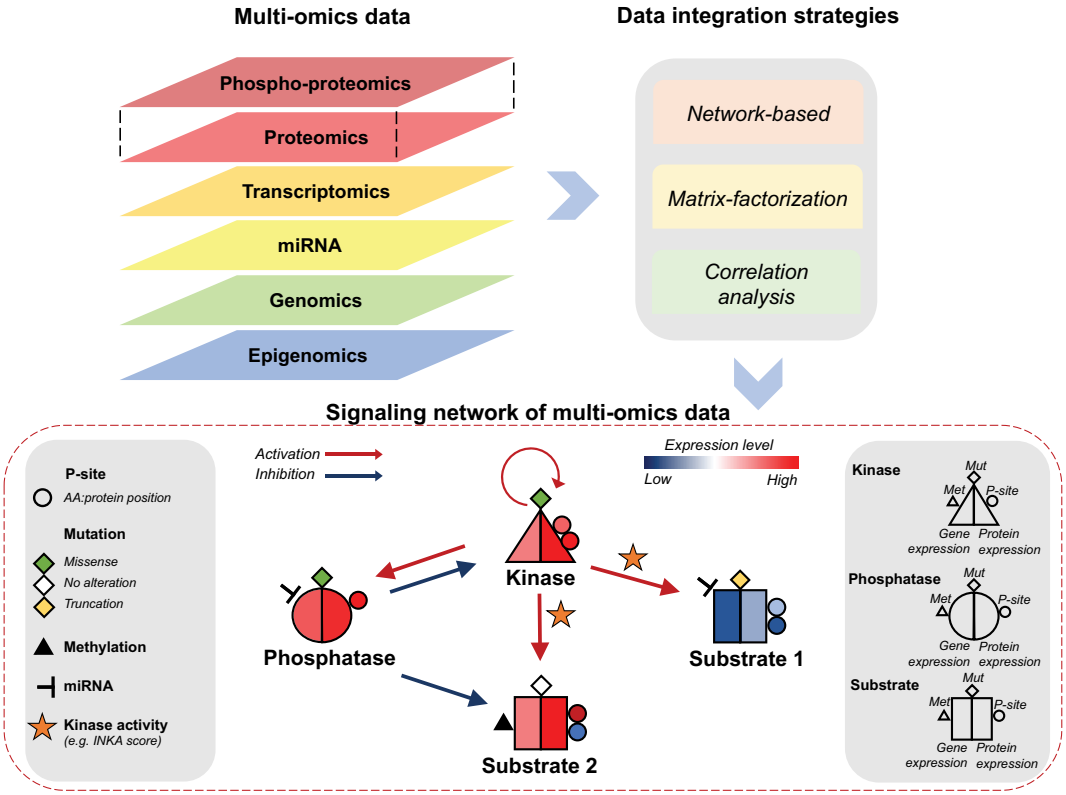


Figure 1. Phosphoproteomics and multi-omics data integration. The upper panel shows different types of –omics data and the main data integration strategies. The lower panel shows an example of signaling network from multi-omics data. Protein kinases are represented by triangles. Phosphatase and substrates are shown in circles and squares, respectively. The left side color of the triangle/circle/square represents the gene expression and the right side color represents the protein expression level. The small circles represent kinase/substrate phosphorylation sites. Green, white, and yellow diamonds represent missense, no alteration, truncation mutation, respectively. Black triangles represent methylation. Red and blue arrows represent activation and inhibitory effect of the kinase/substrate. Star on the arrows indicates kinase activity based on INKA, KARP, or others kinase activity ranking tools.

Landmark cancer molecular profiling programs such as genomics (TCGA)[60] and proteomics (CPTAC,[61] TCPAv3.0[62]) provide multi-omics characterization of individual patients. Combining data from proteomics, phosphoproteomics, and metabolomics with genomics and transcriptomics helps to prioritize disease drivers and channel the attention to specific signaling that can explain the investigated phenotype with the common goal to improve diagnosis, treatment, and cancer prevention. Application studies to the analysis of cancer that include phosphoproteomics are summarized in **Table 2**.

Below we discuss large scale proteogenomics studies focused on multi-omics data integration with phosphoproteomics expanding on specific methods based on network modules and correlation analysis. Moreover, we report on other methods for multi-omics data integration. Those methods are currently not applied to phosphoproteomics and multi-omics data and provide novel opportunities to integrate this data type. Lastly, we mention a list of public multi-omics databases to analyze genomics and proteomics data without any computational knowledge, thus, easily accessible to biologists.

4.1. Proteogenomics

In a typical proteomics study, mass spectrometry data are matched against peptide sequences in a reference proteome database. However, mutated peptides cannot be identified when using a reference proteomics database. This problem can be avoided using an updated database that includes human protein cancer variants[63] or through the integration of multiple data types, collectively called "proteogenomics". Proteogenomics integrates genome, transcriptome, and proteome datasets allowing for correlations of mRNA and protein pairs across samples. Having a general overview of upstream data (genomics and/or

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

Proteomics
Proteomics and Systems Biology
www.proteomics-journal.com

Table 2. Application studies with phosphoproteomics data integration.

Application studies with phosphodata integration							
Data type	Title	Algorithm	Type of cancer	Comments	Phopsho-data level	Type of integration	Reference
CNV, mutation, mRNA, phosphoproteins	Phosphoproteome Integration reveals Patient- Specific Networks in Prostate Cancer	TieDIE	Prostate cancer	The TieDIE algorithm computes a subnetwork of gene and protein interactions. The pathways derived in this way predict interlinking genes that may correspond to essential components of cancer signaling	p-Peptide level collapsed to kinase information with subsequent differential analysis	Network based	[11]
SCNA, mutation, mRNA, proteins, phosphoproteins	Proteogenomic Analysis of Human Colon Cancer Reveals New Therapeutic Opportunities	/	Colon cancer	Proteogenomic integration not only prioritized genomically inferred targets, such as copy-number drivers and mutation-derived neoantigens, but also yielded novel findings. This study demonstrated the utility of proteogenomics in therapeutic hypothesis generation	p-Site level and association with mutational status	/	[12]
Mutation, CNA, mRNA, proteins, phosphoproteins	Integrated Proteogenomic Characterization of HBV- Related Hepatocellular Carcinoma	/	HBV-related hepatocellular carcinoma	joint random forest to join two co-expression networks	p-Protein level		[13]
mRNA, epigenomics, proteins, phosphoproteins	Identification of Therapeutic Targets in Rhabdomyosarcoma through Integrated Genomic, Epigenomic, and Proteomic Analyses	WCGNA	Rhabdomyosarcoma	WGCNA with DE protein and DE phopshosite	p-Site level	ad hoc gene ranking	[14]
DNA, mRNA, proteins, phosphoproteins	Proteogenomic Characterization of Human Early- Onset Gastric Cancer	iClusterPlus, orthogonal non-negative matrix factorization (ONMF), PLS-DA	Gastric cancer	iClusterPlus on DE mRNA, proteins, phosphoproteins, and PLS-DA to identify association of phosphopeptide mutations	MAD score on p-peptide level	matrix-factorization/ cluster modeling	[15]
mRNA, phosphoproteins	Integrated Phosphoproteome and Transcriptome Analysis Reveals Chlamydia-Induced Epithelial-to-Mesenchymal Transition in Host Cells	/	Chlamydia[a)]	p-Peptides in Motif-X for kinase upstream regulators. Kinase-substrate prediction for up- and down-regulated phosphosites	p-Peptide and p-site	/	[99]

(*Continued*)

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

Proteomics
Proteomics and Systems Biology
www.proteomics-journal.com

Table 2. Continued.

Application studies with phosphodata integration							
Data type	Title	Algorithm	Type of cancer	Comments	Phopsho-data level	Type of integration	Reference
WXS and phosphoproteins	Proteogenomic systems analysis identifies targeted therapy resistance mechanisms in EGFR-mutated lung cancer	Dcglob	EGFR-mutated lung cancer	A modified version of the previously published algorithm DCglob	Differential covariances of p-sites	/	[100]
DNA-seq, mRNA, phosphoproteins	Linking Proteomic and Transcriptional Data through the Interactome and Epigenome Reveals a Map of Oncogene-induced Signaling	Prize-collecting Steiner tree (PCST)	GBM	An important aspect of the PCST algorithm is that it is able to naturally account for missing data and false positives. As a result, the final networks are compact, enriched for functionally relevant proteins and the most reliable interactions that include these proteins, and can be used to guide subsequent experiments.	p-Peptide level	Network based	[76]
mRNA, proteins, phosphoproteins	Integrated systems biology analysis of KSHV latent infection reveals viral induction and reliance on peroxisome mediated lipid metabolism	Prize-collecting Steiner tree (PCST)	Kaposi's Sarcoma associated Herpesvirus	This algorithm identifies Steiner nodes, which are proteins that were not implicated in the proteomic or TF analyses but form crucial connections between other important proteins. From OmicsIntegrator R package	p-Peptide level	Network based	[72]
mutation, CNA, mRNA, proteins, phosphoproteins	Proteogenomics connects somatic mutations to signaling in breast cancer	Consensus clustering	Breast cancer	The breadth and depth of proteomic and phosphoproteomic analyses displayed in this study demonstrates the strength of mass-spectrometry-based proteomics	p-Site level	Correlations	[65]
mRNA, proteins, phosphoproteins	Reconstructing targetable pathways in lung cancer by integrating diverse omics data	Abundance score and Prize-Collecting Steiner Tree (PCST)	Non-small cell lung cancer	Using this approach on 13 KRAS-mutant NSCLC cell lines known to be KRAS-Dep or KRAS-Ind, our integrative analysis nominated 115 proteins that were differentially abundant between these two groups (Hochberg-adjusted p-value < 0.05)	p-Protein level	Network based	[80]
Mutation, CNV, mRNA, proteins, phosphoproteins	Proteogenomic integration reveals therapeutic targets in breast cancer xenografts	/	Breast cancer	Collapsed the phosphosites to gene-level phosphorylation values by averaging the phosphosite expressions observed for each gene and converted to z-score	p-Site to gene-level	/	[101]

(*Continued*)

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

Proteomics
Proteomics and Systems Biology
www.proteomics-journal.com

Table 2. Continued.

Application studies with phosphodata integration

Data type	Title	Algorithm	Type of cancer	Comments	Phopsho-data level	Type of integration	Reference
mRNA, proteins, phosphoproteins	Proteomic and phosphoproteomic comparison of human ES and iPS cells	/	Pluripotent stem cells[a)]	Grouped peptides containing multiple sites with other peptides containing the exact same combination of sites. Therefore, we presented a list of phosphorylation isoforms rather than a list of phosphorylated sites	From p-site to p-isoforms	/	[102]
WES, mRNA,circRNA, miRNA, proteins, phosphoproteins	Proteogenomic Characterization of Endometrial Carcinoma	/	Endometrial carcinoma	Kinase activity measure and p-peptide collapsed to gene-level for PTM-outlier	p-Peptide collapsed to gene-level for PTM-outlier	/	[103]
Methylome, mRNA, proteins, phosphoproteins	Aberrant ERBB4-SRC Signaling as a Hallmark of Group 4 Medulloblastoma Revealed by Integrative Phosphoproteomic Profiling	Similarity network fusion (SNF)	Medulloblastoma	This method, described in (Wang et al., 2014), consists of first constructing patient-to-patient similarity networks for each of the available data types and, in a second step, efficiently integrating and fusing them into a final consensus network.	p-Peptide level	Network based	[104]
Methylation, mRNA, proteins, phosphoproteins	Proteomics, Post-translational Modifications, and Integrative Analyses Reveal Molecular Heterogeneity within Medulloblastoma Subgroups	Prize-collecting Steiner tree (PCST)	Medulloblastoma	/	p-Peptide level	Network based	[16]
mRNA, proteins, phosphoproteins	Insights into Impact of DNA Copy Number Alteration and Methylation on the Proteogenomic Landscape of Human Ovarian Cancer via a Multi-omics Integrative Analysis	iProFun	Ovarian cancer	iProFun takes as input the association summary statistics from associating CNAs and methylations of genes to each type of cis-molecular trait, aiming to detect the joint associations of DNA variations and molecular traits in various association patterns	p-Protein level	Multiple linear regression	[17]
Genome, methylome, mRNA, miRNA, proteins, phosphoproteins	A multi-omic analysis of human naïve CD4+ T cells	/	Human naïve CD4+ T cells	/	p-Site level	/	[105]
Proteins, phosphoproteins	Integrating proteomic and phosphoproteomic data for pathway analysis in breast cancer	/	Breast cancer	Integration of proteomics and phosphoproteomics data to perform pathway analysis	p-Protein level	/	[106]

(*Continued*)

Table 2. Continued.

Application studies with phosphodata integration							
Data type	Title	Algorithm	Type of cancer	Comments	Phopsho-data level	Type of integration	Reference
Proteins, phosphoproteins	Integrated analysis of global proteome, phosphoproteome, and glycoproteome enables complementary interpretation of disease-related protein networks	/	Gastric cancer	/	p-Peptide level	/	[107]
mRNA, proteins, phosphoproteins	A System-wide Approach to Monitor Responses to Synergistic BRAF and EGFR Inhibition in Colorectal Cancer Cells	/	Colorectal cancer	/	p-Site level	/	[108]
CNA, mRNA,proteomics, phosphoproteomics	Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer	WCGCNA for subtype clustering and spearman correlation for proteomics and CNA	Ovarian cancer	/	p-Peptide level	/	[66]

[a] Epidemiologically associated with cervical and ovarian cancers.

transcriptomics) and downstream data (proteomics and/or phosphoproteomics) facilitates understanding how post-translational modifications and specific signaling pathways are altered by upstream events such as genetic variants (eQTL), microRNAs, or copy number aberrations. We will now focus on large-scale proteogenomics studies that include phosphoproteomics data and their application methods.

In 2014, Zhang and colleagues[64] performed the first large-scale proteogenomics study analyzing proteomes of 95 colon and rectal tumors previously characterized by TCGA. This analysis demonstrated how proteomics data can enable prioritization of cancer drivers in amplified genomic regions and enable identification of proteomic subtypes, that were only partially overlapping with transcriptomics-based subtypes. These proteomics subtypes also associate differently to clinical outcome. The phosphoproteomic information layer was generated in subsequent cancer proteogenomic studies of the Clinical Proteomic Tumor Analysis Consortium (CPTAC) network. Vasaikar et al. performed a deep molecular characterization of human colon cancer and paired normal adjacent tumors showing that proteogenomic integration may correct inaccurate genomics data-based inferences and lead to unexpected discoveries and therapeutic opportunities.[12] This proteogenomic study not only prioritized genomically inferred targets but also provided novel findings. Phosphoproteomics data associated Rb phosphorylation with increased proliferation and decreased apoptosis suggesting this protein known as tumor suppressor also plays a role as a putative cancer driver.

Other large-scale proteogenomics studies of the CPTAC network that included phosphoproteomics focused on breast, liver, and ovarian cancer.[64–66] Mertins and colleagues described the integration of genomic and (phospho)proteomics information of 105 genomically annotated breast cancers. This analysis also showed the power of proteogenomics to pinpoint novel oncogenic driver genes. First, mRNA and protein/phosphoprotein abundance correlations for each gene/protein pair were analyzed. Identified proteins were mapped to gene names while phosphosites were aggregated to their corresponding protein identifier by calculating the median log-ratio for all sites arising from the protein. Phosphoproteins collapsed to gene names, were subjected to single sample GSEA (ssGSEA) and resulted pathways were clustered by consensus clustering (same approach was used for proteomics). This multi-omics analysis revealed metabolic functions such as amino acid, sugar, and fatty acid metabolism, particularly enriched among positively correlated genes whereas ribosomal, RNA polymerase, and mRNA splicing functions were negatively correlated. An extra analysis was performed for phopshoproteomics: outlier kinase analysis that was carried out to compare the kinase at the level of DNA, RNA, and protein expression within each sample. In this breast cancer study,[65] only the phosphoproteome data shed light on a

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

Proteomics
Proteomics and Systems Biology
www.proteomics-journal.com

G-protein-coupled receptor cluster that was not easily identified from transcriptomics data.

Gao et al. demonstrated that phosphoproteome integration into a multi-omics context of proteomics, transcriptomics, and genomics data revealed alterations of metabolic pathways among the most dramatic differences between tumor and non-tumor hepatitis B virus (HBV)-related liver tissues. These results were given by functional analysis done on the subtype classifications based on transcriptomics, proteomics, and phosphoproteomics data. Again, the data streams were analyzed first separately, then, significance was estimated of the pathway enrichment scores across the datasets using limma statistics to get a single final score for each pathway. Together with metabolic alterations, proliferation and microenvironmental dysregulation signaling stratified patients into three distinct liver cancer subgroups.[13]

Zhang and colleagues presented a proteogenomic integration on 174 high-grade ovarian carcinomas yielding to several insights regarding: i) copy-number alterations that influence the proteome; ii) proteins associated with chromosomal instability; iii) patient stratification by specific protein acetylations. Consensus clustering was applied to compare proteomics subtypes to previous subtypes identified by TCGA. Phosphoproteomics data were mainly used to compare pathway activity between short and long survivors. To avoid the issue of ranking pathways based on phosphopeptides unique to a single pathway and phosphopeptides that map to proteins shared between pathways, Zhang and colleagues performed two analyses: one of all proteins associated with specific pathway and a second analysis in which proteins shared between multiple pathways were excluded. Next, they ranked the pathways based on proteomics, transcriptomics, and CNA information. The results of this parallel analysis of multiple data types showed that RhoA signaling, PDGFRB, integrin-like kinases, NOTCH, and Her2 signaling were more active in patients having short survival. The most significant contribution was given by phosphoproteomics data. Together, these studies highlighted the importance of the phosphoproteomics layer for drug target prediction and association with clinical outcome.

In short, large-scale proteogenomics cancer studies represent an outstanding resource of multi-omics data and exemplify the current state-of-the-art in applied integrative methods. In the next section, we review more in detail the tools used in integrative studies when dealing with phosphoproteomics data that have been applied or not to large scale proteogenomics studies.

4.2. Data-Driven Discovery of Signaling Network Modules

Phosphoproteomics can inform on the state of intracellular phosphorylation both at the pathways level and at the whole cell level. Cascades of protein phosphorylation induced by protein kinases comprise a core mechanism in the integration and propagation of intracellular signals. Hence, a phosphoproteomics layer is valuable in aiding a data-driven discovery of the activation state of signaling network modules.

IOmicsPASS is a tool to search for predictive subnetworks consisting of molecular interactions within and between related omics data types in a supervised analysis setting.[67] Based on user-provided network data and phenotype annotations, iOmicsPASS computes a score for each molecular interaction, and applies a modified nearest shrunken centroid algorithm to the scores to select densely connected subnetworks that can accurately predict each phenotypic group. IOmicsPASS analysis of TCGA/CPTAC with phosphoproteomics breast cancer data[67] highlighted a new transcriptional regulatory network underlying the basal-like subtype as positive protein markers, a result not seen through analysis of individual omics data.

Another network-based approach is Mixed Network Integration (*MiNETi*) developed by Santra et al.[68] on HeLa cell lines. This algorithm operates in four stages. First, it analyzes protein interactome, phosphoproteomics, and transcriptomics data to reconstruct protein-protein interaction (PPI), kinase-substrate (KS), and transcription factor (TF)-DNA interaction networks. These networks are then linked using prior knowledge from PPI databases to construct integrated networks that can be used to track signals emanating from multiprotein complexes and are transmitted via phosphorylation networks to the nucleus, where they modify the activities of transcriptional regulators to control changes in gene expression. *MiNETi* is universally applicable to datasets where PPI changes regulate gene transcription via phosphorylation networks. Santra et al. applied this method to analyze HRAS signaling in HeLa cells from different subcellular compartments, showing that changes in HRAS protein interactions at different sites lead to different kinase activation patterns that differentially regulate gene transcription. The integrated networks provide a topologically and functionally resolved view of HRAS signaling. They reveal distinct HRAS functions including the control of cell migration from the endoplasmic reticulum and TP53-dependent cell survival when signaling from the Golgi apparatus.[68]

OmicsIntegrator is a software package that takes a variety of -omics data as input and identifies putative underlying molecular pathways.[69] The approach applies advanced network optimization algorithms to thousands of molecular interactions to find subnetworks that best explain the data. These subnetworks connect changes observed in gene expression, protein abundance, or other global assays to proteins that may not have been measured in the mass spectrometer due to stochastic sampling, inherent bias, or noise in measurement. The strength of the approach is that it reveals unannotated molecular pathways that would not be detectable by searching pathway databases. The software comprises two individual classifiers, Garnet and Random Forest,[70,71] that can be run together or independently. The tool successfully identified heterogeneity in medulloblastoma subgroups,[16] samples collected from children's hospitals in the United States, highlighting MYC phosphorylation as a prognosis predictor for medulloblastoma subgroup patients. In Sychev et al.[72] a Steiner forest algorithm from the OmicsIntegrator package was used to identify altered cellular pathways in Kaposi's sarcoma associated herpesvirus demonstrating that such pathways were not found when analyzing single biology data stream.

Similarity network fusion (SNF) developed by Wang et al.[73] and applied to genomics data of medulloblastoma samples from Advanced Genomics International Consortium. SNF first constructs a sample similarity network for each of the data types and then iteratively integrates these networks using a novel network fusion method. Working in the sample network space allows SNF to avoid dealing with different scale, collection bias, and noise in different data types. Integrating data

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

Proteomics
Proteomics and Systems Biology
www.proteomics-journal.com

in a non-linear fashion allows SNF to take advantage of the common as well as complementary information in different data types. Cavalli et al. demonstrated that the algorithm was able to find intertumoral heterogeneity within medulloblastoma subgroups using DNA-methylation and gene expression data.[74]

Another method is the prize-collecting Steiner tree problem (PCST).[75] This mathematical model was successfully applied on GBM data by Huang and colleagues[76] to find a network of interactions that link phosphorylation events and differentially transcribed genes. Basically the approach searches for a tree structure in the interactome. The tree structure can be seen as a group of connected nodes (e.g., genes, proteins, miRNAs) and the interactome is usually given a priori using public databases such as STRING,[77] MiNT,[78] and IntACT.[79] As a result, the PCST approach searches for a tree structure in the interactome that connects as many of the experimental data as it can, giving a score (probability). Briefly, in a large gene–gene or protein–protein interaction network, PCST attempts to find sub-networks where many genes are differentially expressed. This method was also applied to address the challenge of data integration in a study conducted by Balbin et al.,[80] where transcriptome, proteome, and phosphoproteome were profiled in a panel of non-small cell lung cancer (NSCLC) cell lines to reconstruct a targetable network associated to KRAS. They first defined an "abundance-score" combining the -omics data to nominate differentially abundant proteins and then used the PCST algorithm to identify functional sub-networks. Finally, LCK was validated as critical gene for cell proliferation in KRAS-dependent NSCLCs and nominated as possible druggable target. Despite the application of this method to multiple cancer datasets, this approach has a limitation of using already pre-defined interaction datasets. While the Steiner tree algorithm was one of the most adopted method in proteogenomic studies integrating PTMs with other -omics, it was always applied to peptide-level phosphoproteomics data, thereby losing important site-specific PTM information.

4.3. Correlation Analysis

Phosphoproteomics provides a regular data matrix that can be incorporated into existing statistical, integrative methods. Multiple-block orthogonal projections to latent structures (OnPLS)[81] is a projection method that simultaneously models multiple data matrices, reducing feature space without relying on a priori biological knowledge and was successfully applied to transcriptomics, proteomics, and metabolomics data showing multi-level oxidative stress response in early onset gastric cancer.[15]

Data clustering plays an important role in integrative -omics analysis, for example, the re-classification of a sample based on its molecular features. iCluster[82] is a penalized latent regression model for joint modeling of multiple types of -omics data. The result is a unique cluster assignment, which includes all –omics data. iCluster achieves dimension reduction of the data and is demonstrated to identify potentially novel subtypes of breast cancer and lung cancer.[83,84] Mo et al. developed iClusterPlus,[85] which is an extension of the iCluster algorithm and incorporates a diversity of data types, including binary mutation status, multicategory copy number states (gain/normal/loss), read count sequencing data, and continuous data, and adopts tailored modeling strategies such as logistic regression, multi-logit regression, Poisson regression, and standard linear regression. The iClusterPlus algorithm has been widely used in subtyping prediction. A recently published example of multi-omics data integration with phosphoproteomics data is the early-onset gastric cancer study from Mun et al.[15] that provided new insights in stratification and cancer-subtype biology from 80 patients. The method for subtyping is summarized in **Figure 2** adapted from publication. First, subtypes were obtained for each individual omics data type (transcriptomics, proteomics, phosphoproteomics, N-glycoproteomics) using orthogonal non-negative matrix factorization (ONMF). The input for ONMF was based on the median absolute deviation (MAD) value using the top 10% or 20% of MADs (depending on the data type). This analysis led to the identification of two RNA subtypes (RNA1-2), four proteomics subtypes (Prot1-4), three phosphoproteomics subtypes (Phos1-3), and three N-glycoproteomics subtypes (Gly1-3). The same input data used for clustering of the individual types of data was used for iClusterPlus (Figure 2A). The resulting clustering identified four different subtypes clearly showing that proteomics and phosphoproteomics data can further subdivide subtypes 1 and 2 where RNA data cannot. Two subtypes were associated with best and worst survival (Figure 2B) and the association of these subtypes to the immune and invasion-related pathways were mainly identified by phosphorylation (circled P) and glycosylation data (circled G) (Figure 2C). The authors concluded that the identification of new subtypes in early-onset gastric cancer was not possible without the proteomics and phosphoproteomics information. Particularly, it would be inefficient to elucidate up-regulation of signaling pathways in the immune and invasive (RHOA) subtypes. Taken together, Mun and colleagues showed how a simple and ready-to-use tool such as iClusterPlus can integrate several data types and extrapolate important features for subtype biology, particularly relevant for patient stratification into drug-sensitive versus resistance. The latter was computed by predicting drug sensitivity based on phosphorylation of ARID1A, CDH1 (immune-subtype), and RHOA (invasive-subtype). However, this publication presents some limitations such as the lack of an in-depth study down to the phosphosite level and a functional validation of the data.

Recently, a new algorithm named iProFun based on multiple linear regression was presented by Song et al.[78] to facilitate screening for cancer drivers and identify proteogenomics functional traits perturbed by DNA copy number alterations (CNAs) and methylations. This tool takes as input the association summary statistics from associating CNAs and methylations of genes to each type of cis-molecular trait, aiming to detect the joint associations of DNA variations and molecular traits in various association patterns. Of particular interest are the genes with "cascading effects" on all cis molecular traits of interest and the genes whose functional regulations are unique at global/phospho protein levels. Downstream enrichment analysis is also embedded into this pipeline. Song et al. demonstrated the feasibility of the algorithm analyzing 570 tumors of ovarian cancer from the TCGA and CPTAC project. The algorithm was able to identify CNAs (e.g., AKT1) driving perturbations on mRNA/proteins levels.

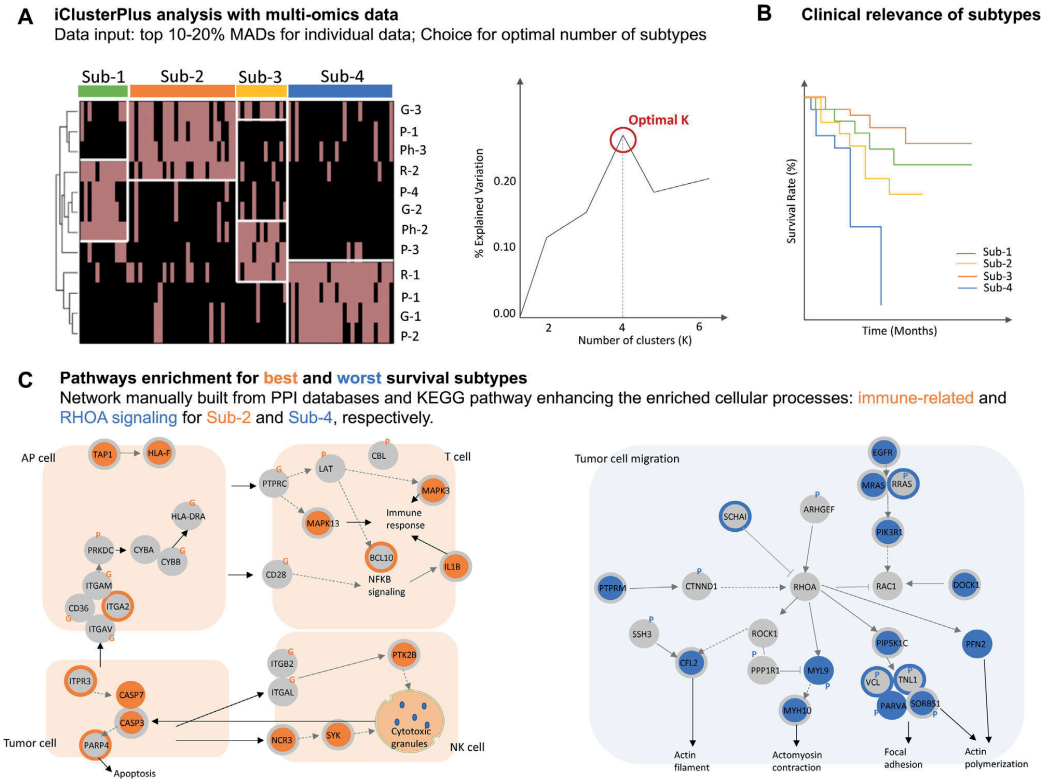


Figure 2. Cancer multi-omics application using the iClusterPlus tool. A) On the left, identification of four subtypes in early-onset gastric cancer from Mun et al. 2019. The clustering shows the importance of different datatypes (clusters pre-defined before) for subtype detection. On the right, percentages in the explained variance with k ranges from 2 to 10 with $k = 4$ the optimal number of subtypes. B) A Kaplan–Meier curve for the predicted subtypes to enhance clinical relevance. C) Signaling networks for the immune subtype (in blue) and the invasive subtype (in dark-green). Networks are manually constructed based on protein–protein interaction (PPI) databases. The center and boundary colors of a node represent whether the corresponding gene and protein were selected as the signatures for the respectively subtypes. Circled P and G of a node indicate that phosphorylated and N-glycosylated peptides from the corresponding protein were selected as subtype's signature. Solid arrows are direct activation and dotted arrows indirect activation.

4.4. Other Opportunities Suitable for Phosphoproteomics Integration

A number of multi-omics integration methods have not been applied yet to phosphoproteomics data. Here, we highlight the most widely used and promising approaches (listed in **Table 3**).

A large number of publications showed the application of weighted gene correlation network analysis (WGCNA)[86] as great approach to find intramodular hub genes as potential biomarkers or drug targets (depending on the research question). By selecting intramodular hubs in consensus modules, WGCNA can also be applied to integrative data mining. Besides its extensive use in integrative biology, there are no publications so far related to the use of WGCNA and PTM datasets. It would be of a great interest to see the application of WGCNA to phosphoproteomics data and at phosphosite level.

Multi-factor analysis (MFA) is devoted to the simultaneous exploration of multiple data tables (-omics data types) where the same individuals are described by several tables of variables. The aims of MFA are similar to those of PCA. Application of MFA to single-cell analysis is done with MOFA (multi-omic factor analysis), a statistical method for integrating multiple modalities of -omics data in an unsupervised fashion.[87] Importantly, MOFA disentangles to what extent each factor is unique to a single data modality or is manifested in multiple modalities, thereby revealing shared axes of variation between the different -omics layers. Once trained, the model output can be used for a range of downstream analyses, including visualization, clustering, and classification of samples in the low-dimensional space(s) spanned by the factors, as well as the automated annotation of factors using (gene set) enrichment analysis, the identification of outlier samples, and the imputation of missing values.

Meng et al. also provided an extension of common gene set enrichment analysis (GSEA) to multi-omics data called MOGSA.[88] This method does not require filtering data to the intersection of features (gene IDs); therefore, all molecular features, including

Table 3. Methods and tools available for general data integration.

Methods/tools/framework available for data integration						
Name	Computational platform	Description	Availability	Number of data types	Type of integration	Reference
WGCNA	R	Integrate 2 datatypes based on module's construction	https://cran.r-project.org/web/packages/WGCNA/index.html	2	Correlation analysis	[86]
MixOmics	R	Offers a wide range of multivariate methods for the exploration and integration of biological datasets with a particular focus on variable selection	https://www.bioconductor.org/packages/release/bioc/html/mixOmics.html	Unlimited	/	[110]
Multiple Dataset integration	Matlab	MDI can integrate information from a wide range of different datasets and data types simultaneously (including the ability to model time series data explicitly using Gaussian processes)	https://warwick.ac.uk/fac/cross_fac/zeeman_institute/zeeman_research/software/	Unlimited	Matrix factorization	[92]
OnPLS	Python	OnPLS extracts a minimal number of globally predictive components that exhibit maximal covariance and correlation	https://github.com/tomlof	Unlimited	Correlation analysis	[111]
MCIA	R	Identifies co-relationships between multiple datasets	https://bioconductor.org/packages/release/bioc/html/omicade4.html	Unlimited	Matrix factorization	[112]
Mogsa	R	Gene set analysis on multiple omics data	https://bioconductor.org/packages/release/bioc/html/mogsa.html	Unlimited	Matrix factorization	[88]
iCluster	R	Joint Gaussian latent variable models	http://www.mskcc.org/mskcc/html/85130.cfm	Unlimited	Matrix factorization	[82]
MFA	R	Combine multiple data sets and to add formalized knowledge	https://cran.r-project.org/web/packages/FactoMineR/index.html	Unlimited	Matrix factorization	
MOFA	R / Python	Method for discovering the principal sources of variation in multi-omics data sets	https://github.com/bioFAM/MOFA	Unlimited	Matrix factorization	[87]
MO network based	R	Network-based analysis of omics with MO optimization	mimomics@itb.cnr.it	Unlimited	Network based	[90]
OmicsIntegrator	Web / Python	Package designed to integrate proteomic data, gene expression data and/or epigenetic data using a protein-protein interaction network	http://fraenkel-nsf.csbi.mit.edu/omicsintegrator/ or https://github.com/fraenkel-lab/OmicsIntegrator	3	Network based	[69]
MiNETi	Matlab / Java / R	Integration of multi-omics datasets that reconstructs and integrates interaction data	https://www.sciencedirect.com/science/article/pii/S2211124719302098?via%3Dihub#mmc8	Unlimited	Network based	[68]
LinkFinder	web	Integration and visualization of two -omics datasets at time	http://www.linkedomics.org/admin.php	2	Correlation analysis	[97]
iProFun	R	Integrative analysis tool to screen for proteogenomic functional traits perturbed by DNA copy number alterations (CNAs) and DNA methylations	https://github.com/songxiaoyu/iProFun	3	Linear regression	[78]

those that lack annotation may be included in the analysis. The publication provides a use case of MOGSA on NCI-60 cell lines panel with transcriptome and proteome data, but can be applied to more than two –omics data types.

A different approach is given by joint and individual variation explained (JIVE) that partitions the variations of the multi-omics data into the sum of three components: 1) a low rank approximation accounting for common variations among -omics data from all platforms, 2) low rank approximations for platform-specific structural variations, and 3) residual noise. JIVE has been extended to joint and individual clustering analysis (JIC) to simultaneously conduct joint and omics-specific clustering on gene expression and miRNA levels on TCGA breast cancer.[89]

Network-based multi-objective (MO) optimization is another method to generate networks of biological components that incorporate multi-omics information.[90] The MO optimization procedure drives the identification of networks that are enriched according to several statistical estimators. The analysis can be applied to different types of -omics and biological interactions. The authors found

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

Proteomics
Proteomics and Systems Biology
www.proteomics-journal.com

protein networks that participate in the establishment of the increased basal differentiation observed in breast tumors of BRCA1-mutation carriers. Additionally, they carried out a network-based comparison among several -omic datasets using transcriptomic data from two types of breast tumors and the corresponding epithelial cells. They found a protein network that shows a strong and coherent (the same direction) differential expression when comparing each tumor with its respective epithelial tissue.

Finally, Bayesian-methods are particularly well suited for integrating different biological data sources, offering a natural setting for integrative analysis including the possibility of incorporating prior information. In contrast, non-Bayesian approaches typically assess different data types in a sequential way or are based on measures of the association between (two) data types rather than on a joint analysis.[91] Multiple dataset integration (MDI)[92] is a Bayesian method for unsupervised integrative modeling of multiple datasets. MDI can integrate information from a wide range of different datasets and data types simultaneously (including the ability to model time series data explicitly using Gaussian processes). Each dataset is modeled using a Dirichlet-multinomial allocation mixture model, with dependencies between these models captured through parameters that describe the agreement among the datasets.

4.5. Open Source Software for Multi-Omics Visualization and Integration

There are open source software platforms that are commonly used for network or pathways visualization. As mentioned before, Cytoscape[54] is the most common used tool where dedicated app such as "Omics Visualizer" is relatively easy-to-use for multi-omics data visualization. Recently, the Cytoscape community has released "omics analysis collection." This app is a collection of already published apps that enables the user to generate a base network from STRING,[77] WikiPathways,[93] and AgilentLiteratureSearch[94] and subsequently projects -omics data on top of them. Afterward, the user is able to perform cluster analysis and functional enrichment while preferred network and heatmap can be visualized as well.

Another extendable pathway analysis toolbox is PathVisio.[95] Unlike Cytoscape, PathVisio allows users to visualize -omics data directly on WikiPathways graphs.[93] Moreover, the user can add/remove specific genes/proteins/metabolites and visualize the specific expression. This tool is intended only for -omics visualization and can be downloaded from https://pathvisio.github.io.

Several databases are emerging as data collection, analysis, and visualization. Databases such as cBioPortal,[96] TCPAv3.0,[62] LinkedOmics,[97] and OmicsDI[98] collect multi-omics data previously characterized by TCGA and CPTAC and provide the ability to visualize/analyze those data together with your own data in a comprehensive way. The web portals are easily accessible via their websites (https://www.cbioportal.org, https://tcpaportal.org, http://www.linkedomics.org, and https://www.omicsdi.org). CBioPortal is the most used website in terms of genomic data integration. The user can upload their own data and run basic analysis like gene–gene correlation, survival analysis, heatmap visualization of different genomics data, and boxplots related to the expression of a certain gene with specific mutation. While cBioPortal, TCPA, and LinkedOmics relies on TCGA and CPTAC sample characterization, OmicsDI is a database storing all data available from several species, tissues, and diseases. The user can look for a specific disease and select the type of data to analyze/download. In addition, cBioPortal provides the option to analyze your own data by installing the instance available on GitHub (https://github.com/cBioPortal/cbioportal). Nevertheless, none of these visualization tools do justice to phosphoproteomics data, as they treat the phosphoproteomics layer as phosphoprotein expression instead of focusing on multiple phosphosites of the protein. At the moment, "Omics Visualizer" is the only Cytoscape plugin available that shows phopshoproteomics data at the phosphosite level together with others -omics data. Here, network connections are based on the STRING database and not yet customizable for different types of input such as kinase-substrate relations.

5. Concluding Remarks and Outlook

Multi-omics analysis has yielded detailed insights into the heterogeneity and complexity of cancers. The main strength of -omics data integration is the capability to track the information flow from upstream (epigenomics, genomics) to downstream effects (proteomics, phosphoproteomics, metabolomics). Recent analyses highlighted the importance of phosphoproteome in understanding tumor complexity, subtypes, and therapeutic vulnerabilities. We have reviewed state-of-the-art methods for analysis of kinase activities, the latest large-scale proteogenomics studies with phosphoproteomics data, and bioinformatics methods where phosphoproteomics data can be treated as another -omics layer. We have highlighted the importance of multi-omics data visualization and data portals to enable the incorporation of expert knowledge in data analysis.

Integration of phosphoproteomics including site-specific peptide information is important. The INKA method for analysis of kinase activities illustrates the power of bringing to the equation in-depth knowledge about phosphosites within kinase activation loops. We expect that research in this direction will benefit from advances in mass spectrometry-based methods for more comprehensive site-specific phosphorylation profiling and the ability to integrate all available phosphorylation datasets for the analysis of individual patients at hand. Developing new tools for -omics data integration remains a constant evolving field. The joint efforts of mathematicians, physicists, bioinformaticians, and biologists will allow cancer researchers to obtain more and more detailed knowledge of tumor complexity, biology, and therapeutic vulnerabilities. We expect these tools will become available for use in the clinic, to enable patient stratification and personalized medicine.

Acknowledgements

The research reported in this publication was supported by the KWF Dutch Cancer Society (Project 2003887, the Netherlands), funded by the KWF grant Dutch Cancer Society (#10212) (C.J.).

Conflict of Interest

The authors declare no conflict of interest.

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

Proteomics
Proteomics and Systems Biology
www.proteomics-journal.com

Keywords

bioinformatics, cancer, data integration, multi-omics, phosphoproteomics

Received: March 24, 2020
Revised: July 9, 2020
Published online: October 8, 2020

[1] D. Hanahan, R. A. Weinberg, *Cell* **2000**, *100*, 57.
[2] K. Rikova, A. Guo, Q. Zeng, A. Possemato, J. Yu, H. Haack, J. Nardone, K. Lee, C. Reeves, Y. Li, Y. Hu, Z. Tan, M. Stokes, L. Sullivan, J. Mitchell, R. Wetzel, J. Macneill, J. M. Ren, J. Yuan, C. E. Bakalarski, J. Villen, J. M. Kornhauser, B. Smith, D. Li, X. Zhou, S. P. Gygi, T. -. L. Gu, R. D. Polakiewicz, J. Rush, M. J. Comb, *Cell* **2007**, *131*, 1190.
[3] M. A. Lemmon, J. Schlessinger, *Cell* **2010**, *141*, 1117.
[4] S. Soverini, M. Mancini, L. Bavaro, M. Cavo, G. Martinelli, *Mol. Cancer* **2018**, *17*, 49.
[5] M. M. Moasser, *Oncogene* **2007**, *26*, 6577.
[6] R. Shackelford, H. El-Osta, *Pharmacogenomics Pers. Med.* **2015**, *8*, 145.
[7] J. G. Paez, *Science* **2004**, *304*, 1497.
[8] R. Dienstmann, L. Vermeulen, J. Guinney, S. Kopetz, S. Tejpar, J. Tabernero, *Nat. Rev. Cancer* **2017**, *17*, 79.
[9] E. E. Voest, R. Bernards, *Cancer Discovery* **2016**, *6*, 130.
[10] W. Du, O. Elemento, *Oncogene* **2015**, *34*, 3215.
[11] J. M. Drake, E. O. Paull, N. A. Graham, J. K. Lee, B. A. Smith, B. Titz, T. Stoyanova, C. M. Faltermeier, V. Uzunangelov, D. E. Carlin, D. T. Fleming, C. K. Wong, Y. Newton, S. Sudha, A. A. Vashisht, J. Huang, J. A. Wohlschlegel, T. G. Graeber, O. N. Witte, J. M. Stuart, *Cell* **2016**, *166*, 1041.
[12] S. Vasaikar, C. Huang, X. Wang, V. A. Petyuk, S. R. Savage, B.o. Wen, Y. Dou, Y. Zhang, Z. Shi, O. A. Arshad, M. A. Gritsenko, L. J. Zimmerman, J. E. Mcdermott, T. R. Clauss, R. J. Moore, R. Zhao, M. E. Monroe, Y.-T. Wang, M. C. Chambers, R. J. C. Slebos, K. S. Lau, Q. Mo, L. Ding, M. Ellis, M. Thiagarajan, C. R. Kinsinger, H. Rodriguez, R. D. Smith, K. D. Rodland, D. C. Liebler, et al., *Cell* **2019**, *177*, 1035.
[13] Q. Gao, H. Zhu, L. Dong, W. Shi, R. Chen, Z. Song, C. Huang, J. Li, X. Dong, Y. Zhou, Q. Liu, L. Ma, X. Wang, J. Zhou, Y. Liu, E. Boja, A. I. Robles, W. Ma, P. Wang, Y. Li, L.i. Ding, B.o. Wen, B. Zhang, H. Rodriguez, D. Gao, H.u. Zhou, J. Fan, *Cell* **2019**, *179*, 561.
[14] E. Stewart, J. Mcevoy, H. Wang, X. Chen, V. Honnell, M. Ocarz, B. Gordon, J. Dapper, K. Blankenship, Y. Yang, Y. Li, T. I. Shaw, J.-H. Cho, X. Wang, B. Xu, P. Gupta, Y. Fan, Y. Liu, M. Rusch, L. Griffiths, J. Jeon, B. B. Freeman, M. R. Clay, A. Pappo, J. Easton, S. Shurtleff, A. Shelat, X. Zhou, K. Boggs, H. Mulder, et al., *Cancer Cell* **2018**, *34*, 411.
[15] D.-G. Mun, J. Bhin, S. Kim, H. Kim, J. H. Jung, Y. Jung, Y. E. Jang, J. M. Park, H. Kim, Y. Jung, H. Lee, J. Bae, S. Back, S.-J. Kim, J. Kim, H. Park, H. Li, K.-B. Hwang, Y. S. Park, J. H. Yook, B. S. Kim, S. Y. Kwon, S. W. Ryu, D. Y. Park, T. Y. Jeon, D. H. Kim, J.-H. Lee, S.-U. Han, K. S. Song, D. Park, et al., *Cancer Cell* **2019**, *35*, 111.
[16] T. C. Archer, T. Ehrenberger, F. Mundt, M. P. Gold, K. Krug, C. K. Mah, E. L. Mahoney, C. J. Daniel, A. Lenail, D. Ramamoorthy, P. Mertins, D. R. Mani, H. Zhang, M. A. Gillette, K. Clauser, M. Noble, L. C. Tang, J. Pierre-François, J. Silterra, J. Jensen, P. Tamayo, A. Korshunov, S. M. Pfister, M. Kool, P. A. Northcott, R. C. Sears, J. O. Lipton, S. A. Carr, J. P. Mesirov, S. L. Pomeroy, et al., *Cancer Cell* **2018**, *34*, 396.
[17] X. Song, J. Ji, K. J. Gleason, F. Yang, J. A. Martignetti, L. S. Chen, P. Wang, *Mol. Cell. Proteomics* **2019**, *18*, S52.
[18] T. S. Batth, M. Papetti, A. Pfeiffer, M. A. X. Tollenaere, C. Francavilla, J. V. Olsen, *Cell Rep.* **2018**, *22*, 2784.
[19] J. R. Yates, in *Proteomics for Biological Discovery* (Eds: T. D. Veenstra, J. R. Yates), John Wiley & Sons, New York **2019**, p. 291.
[20] A. Hogrebe, L. Von Stechow, D. B. Bekker-Jensen, B. T. Weinert, C. D. Kelstrup, J. V. Olsen, *Nat. Commun.* **2018**, *9*, 1045.
[21] P. V. Hornbeck, B. Zhang, B. Murray, J. M. Kornhauser, V. Latham, E. Skrzypek, *Nucleic Acids Res.* **2015**, *43*, D512.
[22] G. Manning, *Science* **2002**, *298*, 1912.
[23] G. Duan, X. Li, M. Köhn, *Nucleic Acids Res.* **2015**, *43*, D531.
[24] M. Kim, D.-S. Shin, J. Kim, Y.-S. Lee, *Biopolymers* **2010**, *94*, 753.
[25] T. Hunter, *Cell* **1995**, *80*, 225.
[26] I. Song, Y. J. Jeong, J. Park, S. Shim, S. Jang, *Sci. Rep.* **2017**, *7*, 11797.
[27] H. Horn, E. M. Schoof, J. Kim, X. Robin, M. L. Miller, F. Diella, A. Palma, G. Cesareni, L. J. Jensen, R. Linding, *Nat. Methods* **2014**, *11*, 603.
[28] H. Dinkel, C. Chica, A. Via, C. M. Gould, L. J. Jensen, T. J. Gibson, F. Diella, *Nucleic Acids Res.* **2011**, *39*, D261.
[29] C. Terfve, T. Cokelaer, D. Henriques, A. Macnamara, E. Goncalves, M. K. Morris, M. V. Iersel, D. A. Lauffenburger, J. Saez-Rodriguez, *BMC Syst. Biol.* **2012**, *6*, 133.
[30] M. W. H. Pinkse, P. M. Uitto, M. J. Hilhorst, B. Ooms, A. J. R. Heck, *Anal. Chem.* **2004**, *76*, 3935.
[31] M. R. Larsen, T. E. Thingholm, O. N. Jensen, P. Roepstorff, T. J. D. Jørgensen, *Mol. Cell. Proteomics* **2005**, *4*, 873.
[32] N. Sugiyama, T. Masuda, K. Shinoda, A. Nakamura, M. Tomita, Y. Ishihama, *Mol. Cell. Proteomics* **2007**, *6*, 1103.
[33] L. Zhou, R. W. Beuerman, C. M. Chan, S. Z. Zhao, X. R. Li, H. Yang, L. Tong, S. Liu, M. E. Stern, D. Tan, *J. Proteome Res.* **2009**, *8*, 4889.
[34] L. Andersson, J. Porath, *Anal. Biochem.* **1986**, *154*, 250.
[35] J. Rush, A. Moritz, K. A. Lee, A. Guo, V. L. Goss, E. J. Spek, H. Zhang, X.-M. Zha, R. D. Polakiewicz, M. J. Comb, *Nat. Biotechnol.* **2005**, *23*, 94.
[36] C. M. Potel, S. Lemeer, A. J. R. Heck, *Anal. Chem.* **2019**, *91*, 126.
[37] J. Cox, M. Mann, *Nat. Biotechnol.* **2008**, *26*, 1367.
[38] D. N. Perkins, D. J. C. Pappin, D. M. Creasy, J. S. Cottrell, *Electrophoresis* **1999**, *20*, 3551.
[39] S. A. Beausoleil, J. Villén, S. A. Gerber, J. Rush, S. P. Gygi, *Nat. Biotechnol.* **2006**, *24*, 1285.
[40] J.ü. Cox, N. Neuhauser, A. Michalski, R. A. Scheltema, J. V. Olsen, M. Mann, *J. Proteome Res.* **2011**, *10*, 1794.
[41] T. Taus, T. Köcher, P. Pichler, C. Paschke, A. Schmidt, C. Henrich, K. Mechtler, *J. Proteome Res.* **2011**, *10*, 5354.
[42] M. M. Savitski, S. Lemeer, M. Boesche, M. Lang, T. Mathieson, M. Bantscheff, B. Kuster, *Mol. Cell. Proteomics* **2011**, *10*, M110.003830.
[43] D. B. Bekker-Jensen, O. M. Bernhardt, A. Hogrebe, A. Martinez-Val, L. Verbeke, T. Gandhi, C. D. Kelstrup, L. Reiter, J. V. Olsen, *Nat. Commun.* **2020**, *11*, 787.
[44] A. Lachmann, A. Ma'ayan, *Bioinformatics* **2009**, *25*, 684.
[45] D. D. Wiredja, M. Koyutürk, M. R. Chance, *Bioinformatics* **2017**, *33*, 3489.
[46] P. Casado, J.-C. Rodriguez-Prados, S. C. Cosulich, S. Guichard, B. Vanhaesebroeck, S. Joel, P. R. Cutillas, *Sci. Signaling* **2013**, *6*, rs6.
[47] S.-B. Suo, J.-D. Qiu, S.-P. Shi, X. Chen, R.-P. Liang, *Sci. Rep.* **2015**, *4*, 4524.
[48] C. Weidner, C. Fischer, S. Sauer, *Bioinformatics* **2014**, *30*, 3410.
[49] M. Mischnik, F. Sacco, J. Cox, H.-C. Schneider, M. Schäfer, M. Hendlich, D. Crowther, M. Mann, T. Klabunde, *Bioinformatics* **2016**, *32*, 424.
[50] K. Krug, P. Mertins, B. Zhang, P. Hornbeck, R. Raju, R. Ahmad, M. Szucs, F. Mundt, D. Forestier, J. Jane-Valbuena, H. Keshishian, M. A. Gillette, P. Tamayo, J. P. Mesirov, J. D. Jaffe, S. A. Carr, D. R. Mani, *Mol. Cell. Proteomics* **2019**, *18*, 576.
[51] R. Beekhof, C. Alphen, A. A. Henneman, J. C. Knol, T. V. Pham, F. Rolfs, M. Labots, E. Henneberry, T. Y. Le Large, R. R. Haas, S. R.

ADVANCED SCIENCE NEWS
www.advancedsciencenews.com

Proteomics
Proteomics and Systems Biology
www.proteomics-journal.com

Piersma, V. Vurchio, A. Bertotti, L. Trusolino, H. M. Verheul, C. R. Jimenez, *Mol. Sys. Biol.* **2019**, *15*, e8250.
[52] C. van Alphen, J. Cloos, R. Beekhof, D. G. J. Cucchi, S. R. Piersma, J. C. Knol, A. A. Henneman, T. V. Pham, J. van Meerloo, G. J. Ossenkoppele, H. M. W. Verheul, J. J. W. M. Janssen, C. R. Jimenez, *Mol. Cell. Proteomics* **2020**, *19*, 884.
[53] D. Ochoa, A. F. Jarnuczak, M. Gehre, M. Soucheray, A. A. Kleefeldt, C. Viéitez, A. Hill, L. Garcia-Alonso, D. L. Swaney, J. A. Vizcaìno, K. Noh, P. Beltrao, *Nat. Biotech* **2020**, *38*, 365.
[54] P. Shannon, *Genome Res.* **2003**, *13*, 2498.
[55] S. Lotia, J. Montojo, Y. Dong, G. D. Bader, A. R. Pico, *Bioinformatics* **2013**, *29*, 1350.
[56] M. Legeay, N. T. Doncheva, J. H. Morris, L. J. Jensen, *F1000Research* **2020**, *9*, 157.
[57] L. M. Raaijmakers, P. Giansanti, P. A. Possik, J. Mueller, D. S. Peeper, A. J. R. Heck, A. F. M. Altelaar, *J. Proteome Res.* **2015**, *14*, 4332.
[58] Y. Narushima, H. Kozuka-Hata, K. Tsumoto, J.-I. Inoue, M. Oyama, *Bioinformatics* **2016**, *32*, 2083.
[59] A. P. Tay, C. N. I. Pang, D. L. Winter, M. R. Wilkins, *J. Proteome Res.* **2017**, *16*, 1988.
[60] The Cancer Genome Atlas Research Network, J. N. Weinstein, E. A. Collisson, G. B. Mills, K. R. Mills Shaw, B. A. Ozenberger, K. Ellrott, I. Shmulevich, C. Sander, J. M. Stuart, *Nat. Genet.* **2013**, *45*, 1113.
[61] P. A. Rudnick, S. P. Markey, J. Roth, Y. Mirokhin, X. Yan, D. V. Tchekhovskoi, N. J. Edwards, R. R. Thangudu, K. A. Ketchum, C. R. Kinsinger, M. Mesri, H. Rodriguez, S. E. Stein, *J. Proteome Res.* **2016**, *15*, 1023.
[62] M.-J. M. Chen, B. Searle, L. Pino, D. Plubell, D. Faivre, G. Merrihew, J. Egertson, S. Ting, B. MacLean, *Mol. Cell. Proteomics* **2019**, *18*, S15.
[63] M. Zhang, B. Wang, J. Xu, X. Wang, L. Xie, B. Zhang, Y. Li, J. Li, *J. Proteome Res.* **2017**, *16*, 421.
[64] B. Zhang, J. Wang, X. Wang, J. Zhu, Q. Liu, Z. Shi, M. C. Chambers, L. J. Zimmerman, K. F. Shaddox, S. Kim, S. R. Davies, S. Wang, P. Wang, C. R. Kinsinger, R. C. Rivers, H. Rodriguez, R. Reid Townsend, M. J. C. Ellis, S. A. Carr, D. L. Tabb, R. J. Coffey, R. J. C. Slebos, D. C. Liebler, NCI CPTAC, *Nature* **2014**, *513*, 382.
[65] P. Mertins, D. R. Mani, K. V. Ruggles, M. A. Gillette, K. R. Clauser, P. Wang, X. Wang, J. W. Qiao, S. Cao, F. Petralia, E. Kawaler, F. Mundt, K. Krug, Z. Tu, J. T. Lei, M. L. Gatza, M. Wilkerson, C. M. Perou, V. Yellapantula, K.-l. Huang, C. Lin, M. D. McLellan, P. Yan, S. R. Davies, R. Reid Townsend, S. J. Skates, J. Wang, B. Zhang, C. R. Kinsinger, M. Mesri, H. Rodriguez, L. Ding, A. G. Paulovich, D. Fenyö, M. J. Ellis, S. A. Carr, NCI CPTAC, *Nature* **2016**, *534*, 55.
[66] H. Zhang, T. Liu, Z. Zhang, S. H. Payne, B. Zhang, J. E. Mcdermott, J.-Y. Zhou, V. A. Petyuk, L. Chen, D. Ray, S. Sun, F. Yang, L. Chen, J. Wang, P. Shah, S. W. Cha, P. Aiyetan, S. Woo, Y. Tian, M. A. Gritsenko, T. R. Clauss, C. Choi, M. E. Monroe, S. Thomas, S. Nie, C. Wu, R. J. Moore, K.-H. Yu, D. L. Tabb, D. Fenyö, et al., *Cell* **2016**, *166*, 755.
[67] H. W. L. Koh, D. Fermin, C. Vogel, K. P. Choi, R. M. Ewing, H. Choi, *npj Sys. Biol. Appl.* **2019**, *5*, 22.
[68] T. Santra, A. Herrero, J. Rodriguez, A. Von Kriegsheim, L. F. Iglesias-Martinez, T. Schwarzl, D. Higgins, T.-T. Aye, A. J. R. Heck, F. Calvo, L. Agudo-Ibáñez, P. Crespo, D. Matallanas, W. Kolch, *Cell Rep.* **2019**, *26*, 3100.
[69] N. Tuncbag, S. J. C. Gosline, A. Kedaigle, A. R. Soltis, A. Gitter, E. Fraenkel, *PLoS Comput. Biol.* **2016**, *12*, e1004879.
[70] G. X. Y. Zheng, J. M. Terry, P. Belgrader, P. Ryvkin, Z. W. Bent, R. Wilson, S. B. Ziraldo, T. D. Wheeler, G. P. McDermott, J. Zhu, M. T. Gregory, J. Shuga, L. Montesclaros, J. G. Underwood, D. A. Masquelier, S. Y. Nishimura, M. Schnall-Levin, P. W. Wyatt, C. M. Hindson, R. Bharadwaj, A. Wong, K. D. Ness, L. W. Beppu, H. Joachim Deeg, C. McFarland, K. R. Loeb, W. J. Valente, N. G. Ericson, E. A. Stevens, J. P. Radich, T. S. Mikkelsen, B. J. Hindson, J. H. Bielas, *Nat. Commun.* **2017**, *8*, 14049.
[71] T. K. Ho, *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 832.
[72] Z. E. Sychev, A. Hu, T. A. Dimaio, A. Gitter, N. D. Camp, W. S. Noble, A. Wolf-Yadlin, M. Lagunoff, *PLoS Pathog.* **2017**, *13*, e1006256.
[73] B.o. Wang, A. M. Mezlini, F. Demir, M. Fiume, Z. Tu, M. Brudno, B. Haibe-Kains, A. Goldenberg, *Nat. Methods* **2014**, *11*, 333.
[74] F. M. G. Cavalli, M. Remke, L. Rampasek, J. Peacock, D. J. H. Shih, B. Luu, L. Garzia, J. Torchia, C. Nor, A. S. Morrissy, S. Agnihotri, Y. Y. Thompson, C. M. Kuzan-Fischer, H. Farooq, K. Isaev, C. Daniels, B.-K. Cho, S.-K. Kim, K.-C. Wang, J. Y. Lee, W. A. Grajkowska, M. Perek-Polnik, A. Vasiljevic, C. Faure-Conter, A. Jouvet, C. Giannini, A. A. Nageswara Rao, K. K. W. Li, H.-K. Ng, C. G. Eberhart, et al., *Cancer Cell* **2017**, *31*, 737.
[75] I. Ljubić, R. Weiskircher, U. Pferschy, G. W. Klau, P. Mutzel, M. Fischetti, *Math. Program.* **2006**, *105*, 427.
[76] S.-S. C. Huang, D. C. Clarke, S. J. C. Gosline, A. Labadorf, C. R. Chouinard, W. Gordon, D. A. Lauffenburger, E. Fraenkel, *PLoS Comput. Biol.* **2013**, *9*, e1002887.
[77] B. Snel, *Nucleic Acids Res.* **2000**, *28*, 3442.
[78] L. Licata, L. Briganti, D. Peluso, L. Perfetto, M. Iannuccelli, E. Galeota, F. Sacco, A. Palma, A. P. Nardozza, E. Santonico, L. Castagnoli, G. Cesareni, *Nucleic Acids Res.* **2012**, *40*, D857.
[79] S. Orchard, M. Ammari, B. Aranda, L. Breuza, L. Briganti, F. Broackes-Carter, N. H. Campbell, G. Chavali, C. Chen, N. Del-Toro, M. Duesbury, M. Dumousseau, E. Galeota, U. Hinz, M. Iannuccelli, S. Jagannathan, R. Jimenez, J. Khadake, A. Lagreid, L. Licata, R. C. Lovering, B. Meldal, A. N. Melidoni, M. Milagros, D. Peluso, L. Perfetto, P. Porras, A. Raghunath, S. Ricard-Blum, B. Roechert, et al., *Nucleic Acids Res.* **2014**, *42*, D358.
[80] O. A. Balbin, J. R. Prensner, A. Sahu, A. Yocum, S. Shankar, R. Malik, D. Fermin, S. M. Dhanasekaran, B. Chandler, D. Thomas, D. G. Beer, X. Cao, A. I. Nesvizhskii, A. M. Chinnaiyan, *Nat. Commun.* **2013**, *4*.
[81] V. Srivastava, O. Obudulu, J. Bygdell, T. Löfstedt, P. Rydén, R. Nilsson, M. Ahnlund, A. Johansson, P. Jonsson, E. Freyhult, J. Qvarnström, J. Karlsson, M. Melzer, T. Moritz, J. Trygg, T. R. Hvidsten, G. Wingsle, *BMC Genomics* **2013**, *14*, 893.
[82] R. Shen, A. B. Olshen, M. Ladanyi, *Bioinformatics* **2009**, *25*, 2906.
[83] J. R. Pollack, T. Sorlie, C. M. Perou, C. A. Rees, S. S. Jeffrey, P. E. Lonning, R. Tibshirani, D. Botstein, A.-L. Borresen-Dale, P. O. Brown, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12963.
[84] D. Chitale, Y. Gong, B. S. Taylor, S. Broderick, C. Brennan, R. Somwar, B. Golas, L. Wang, N. Motoi, J. Szoke, J. M. Reinersman, J. Major, C. Sander, V. E. Seshan, M. F. Zakowski, V. Rusch, W. Pao, W. Gerald, M. Ladanyi, *Oncogene* **2009**, *28*, 2773.
[85] Q. Mo, S. Wang, V. E. Seshan, A. B. Olshen, N. Schultz, C. Sander, R. S. Powers, M. Ladanyi, R. Shen, *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 4245.
[86] P. Langfelder, S. Horvath, *BMC Bioinf.* **2008**, *9*, 559.
[87] R. Argelaguet, B. Velten, D. Arnol, S. Dietrich, T. Zenz, J. C. Marioni, F. Buettner, W. Huber, O. Stegle, *Mol. Sys. Biol.* **2018**, *14*, e8124.
[88] C. Meng, A. Basunia, B. Peters, A. Moghaddas Gholami, B. Kuster, A. C. Culhane, *Mol. Cell. Proteomics* **2019**, *18*, S153.
[89] K. H. Hellton, M. Thoresen, *Biostatistics* **2016**, *17*, 537.
[90] E. Mosca, L. Milanesi, *Mol. BioSyst.* **2013**, *9*, 2971.
[91] K. Ickstadt, M. Schäfer, M. Zucknick, *Ann. Rev. Stat. Its Appl.* **2018**, *5*, 141.
[92] P. Kirk, J. E. Griffin, R. S. Savage, Z. Ghahramani, D. L. Wild, *Bioinformatics* **2012**, *28*, 3290.
[93] D. N. Slenter, M. Kutmon, K. Hanspers, A. Riutta, J. Windsor, N. Nunes, J. Mélius, E. Cirillo, S. L. Coort, D. Digles, F. Ehrhart, P. Giesbertz, M. Kalafati, M. Martens, R. Miller, K. Nishida, L. Rieswijk, A.

Waagmeester, L. M. T. Eijssen, C. T. Evelo, A. R. Pico, E. L. Willighagen, *Nucleic Acids Res.* **2018**, *46*, D661.
[94] M. S. Cline, M. Smoot, E. Cerami, A. Kuchinsky, N. Landys, C. Workman, R. Christmas, I. Avila-Campilo, M. Creech, B. Gross, K. Hanspers, R. Isserlin, R. Kelley, S. Killcoyne, S. Lotia, S. Maere, J. Morris, K. Ono, V. Pavlovic, A. R. Pico, A. Vailaya, P.-L. Wang, A. Adler, B. R. Conklin, L. Hood, M. Kuiper, C. Sander, I. Schmulevich, B. Schwikowski, G. J. Warner, T. Ideker, G. D. Bader, *Nat. Protoc.* **2007**, *2*, 2366.
[95] M. Kutmon, M. P. Van Iersel, A. Bohler, T. Kelder, N. Nunes, A. R. Pico, C. T. Evelo, *PLoS Comput. Biol.* **2015**, *11*, e1004085.
[96] E. Cerami, J. Gao, U. Dogrusoz, B. E. Gross, S. O. Sumer, B. A. Aksoy, A. Jacobsen, C. J. Byrne, M. L. Heuer, E. Larsson, Y. Antipin, B. Reva, A. P. Goldberg, C. Sander, N. Schultz, *Cancer Discovery* **2012**, *2*, 401.
[97] S. V. Vasaikar, P. Straub, J. Wang, B. Zhang, *Nucleic Acids Res.* **2018**, *46*, D956.
[98] Y. Perez-Riverol, M. Bai, F. Da Veiga Leprevost, S. Squizzato, Y. M.i. Park, K. Haug, A. J. Carroll, D. Spalding, J. Paschall, M. Wang, N. Del-Toro, T. Ternent, P. Zhang, N. Buso, N. Bandeira, E. W. Deutsch, D. S. Campbell, R. C. Beavis, R. M. Salek, U. Sarkans, R. Petryszak, M. Keays, E. Fahy, M. Sud, S. Subramaniam, A. Barbera, R. C. Jiménez, A. I. Nesvizhskii, S.-A. Sansone, C. Steinbeck, et al., *Nat. Biotechnol.* **2017**, *35*, 406.
[99] P. K. Zadora, C. Chumduri, K. Imami, H. Berger, Y. Mi, M. Selbach, T. F. Meyer, R. K. Gurumurthy, *Cell Rep.* **2019**, *26*, 1286.
[100] D. Treue, M. Bockmayr, A. Stenzinger, D. Heim, S. Hester, F. Klauschen, *Int. J. Cancer* **2019**, *144*, 545.
[101] G. Huang, M. Yuan, M. Chen, L. Li, W. You, H. Li, J. J. Cai, G. Ji, *Analyst* **2017**, *142*, 3588.
[102] D. H. Phanstiel, J. Brumbaugh, C. D. Wenger, S. Tian, M. D. Probasco, D. J. Bailey, D. L. Swaney, M. A. Tervo, J. M. Bolin, V. Ruotti, R. Stewart, J. A. Thomson, J. J. Coon, *Nat. Methods* **2011**, *8*, 821.
[103] Y. Dou, E. A. Kawaler, D. Cui Zhou, M. A. Gritsenko, C. Huang, L. Blumenberg, A. Karpova, V. A. Petyuk, S. R. Savage, S. Satpathy, W. Liu, Y. Wu, C. F. Tsai, B. Wen, Z. Li, S. Cao, J. Moon, Z. Shi, M. Cornwell, M. A. Wyczalkowski, R. K. Chu, S. Vasaikar, H. Zhou, Q. Gao, R. J. Moore, K. Li, S. Sethuraman, M. E. Monroe, R. Zhao, D. Heiman, et al., *Cell* **2020**, *180*, 729.
[104] A. Forget, L. Martignetti, S. Puget, L. Calzone, S. Brabetz, D. Picard, A. Montagud, S. Liva, A. Sta, F. Dingli, G. Arras, J. Rivera, D. Loew, A. Besnard, J. Lacombe, M. Pagès, P. Varlet, C. Dufour, H. Yu, A. L. Mercier, E. Indersie, A. Chivet, S. Leboucher, L. Sieber, K. Beccaria, M. Gombert, F. D. Meyer, N. Qin, J. Bartl, L. Chavez, et al., *Cancer Cell* **2018**, *34*, 379.
[105] C. J. Mitchell, D. Getnet, M.-S. Kim, S. S. Manda, P. Kumar, T.-C. Huang, S. M. Pinto, R. S. Nirujogi, M. Iwasaki, P. G. Shaw, X. Wu, J. Zhong, R. Chaerkady, A. Marimuthu, B. Muthusamy, N. A. Sahasrabuddhe, R. Raju, C. Bowman, L. Danilova, J. Cutler, D. S. Kelkar, C. G. Drake, T. S. K. Prasad, L. Marchionni, P. N. Murakami, A. F. Scott, L. Shi, J. Thierry-Mieg, D. Thierry-Mieg, R. Irizarry, et al., *BMC Syst. Biol.* **2015**, *9*.
[106] J. Ren, B. Wang, J. Li, *BMC Syst. Biol.* **2018**, *12*.
[107] J.-M. Park, *et al.*, *Sci. Rep.* **2015**, *5*.
[108] A. Ressa, *et al.*, *Mol. Cell. Proteomics* **2018**, *17*, 1892.
[109] F. Rohart, B. Gautier, A. Singh, K.-A. Cao, *PLoS Comput. Biol.* **2017**, *13*, e1005752.
[110] T. Löfstedt, J. Trygg, *J. Chemom.* **2011**, n/a.
[111] C. Meng, B. Kuster, A. C. Culhane, A. Gholami, *BMC Bioinformatics* **2014**, *15*, 162.
[112] K. Yu, Q. Zhang, Z. Liu, Q. Zhao, X. Zhang, Y. Wang, Z.-X. Wang, Y. Jin, X. Li, Z.-X. Liu, R.-H. Xu, *Nucleic Acids Res.* **2019**, *47*, D451.
[113] S. Ullah, S. Lin, Y. Xu, W. Deng, L. Ma, Y. Zhang, Z. Liu, Y. Xue, *Sci. Rep.* **2016**, *6*.
[114] D. J. B. Clarke, M. V. Kuleshov, B. M. Schilder, D. Torre, M. E. Duffy, A. B. Keenan, A. Lachmann, A. S. Feldmann, G. W. Gundersen, M. C. Silverstein, Z. Wang, A. Ma'ayan, *Nucleic Acids Res.* **2018**, *46*, W171.
[115] C. N. Schlaffner, G. J. Pirklbauer, A. Bender, J. S. Choudhary, *Cell Syst.* **2017**, *5*, 152.

Giulia Mantini is a Ph.D. student in computational biology at the Amsterdam UMC–Cancer Center. Her Ph.D. project involves the identification of drug targets and diagnostic markers in solid and liquid tumor biopsies with a focus on cancer proteomics and phosphoproteomics, data integration, systems biology, and machine learning. She holds a master's degree in bioinformatics from Alma Mater Studiorum Bologna, Italy, while investigating a new methodology for massive parallel sequencing and barcoding in a joint project with the SciLifeLab–Stockholm in the Molecular Diagnostic Group. She is currently in the last year of her Ph.D.

Chapter 3

Co-expression analysis of pancreatic cancer proteome reveals biology and prognostic biomarkers

Mantini G, Vallés AM*, Le Large TYS*, Capula M, Funel N, Pham TV, Piersma SR, Kazemier G, Bijlsma MF, Giovannetti E*, Jimenéz CR*
*These authors contributed equally

Cellular Oncology, 2020

Cellular Oncology (2020) 43:1147–1159
https://doi.org/10.1007/s13402-020-00548-y

ORIGINAL PAPER

Co-expression analysis of pancreatic cancer proteome reveals biology and prognostic biomarkers

G. Mantini[1,2] · A. M. Vallés[1] · T. Y. S. Le Large[1,3,4] · M. Capula[2] · N. Funel[5] · T. V. Pham[1] · S. R. Piersma[1] · G. Kazemier[4] · M. F. Bijlsma[5,6] · E. Giovannetti[1,2] · C. R. Jimenez[1]

Accepted: 30 June 2020 / Published online: 29 August 2020

Abstract

Purpose Despite extensive biological and clinical studies, including comprehensive genomic and transcriptomic profiling efforts, pancreatic ductal adenocarcinoma (PDAC) remains a devastating disease, with a poor survival and limited therapeutic options. The goal of this study was to assess co-expressed PDAC proteins and their associations with biological pathways and clinical parameters.

Methods Correlation network analysis is emerging as a powerful approach to infer tumor biology from omics data and to prioritize candidate genes as biomarkers or drug targets. In this study, we applied a weighted gene co-expression network analysis (WGCNA) to the proteome of 20 surgically resected PDAC specimens (PXD015744) and confirmed its clinical value in 82 independent primary cases.

Results Using WGCNA, we obtained twelve co-expressed clusters with a distinct biology. Notably, we found that one module enriched for metabolic processes and epithelial-mesenchymal-transition (EMT) was significantly associated with overall survival ($p = 0.01$) and disease-free survival ($p = 0.03$). The prognostic value of three proteins (SPTBN1, KHSRP and PYGL) belonging to this module was confirmed using immunohistochemistry in a cohort of 82 independent resected patients. Risk score evaluation of the prognostic signature confirmed its association with overall survival in multivariate analyses. Finally, immunofluorescence analysis confirmed co-expression of SPTBN1 and KHSRP in Hs766t PDAC cells.

Conclusions Our WGCNA analysis revealed a PDAC module enriched for metabolic and EMT-associated processes. In addition, we found that three of the proteins involved were associated with PDAC survival.

Keywords Pancreatic cancer · Protein co-expression · Systems biology · Proteomics · WGCNA · Prognostic biomarkers

A. M. Vallés and T. Y. S. Le Large contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s13402-020-00548-y) contains supplementary material, which is available to authorized users.

✉ E. Giovannetti
e.giovannetti@amsterdamumc.nl

✉ C. R. Jimenez
c.jimenez@amsterdamumc.nl

[1] Amsterdam UMC, Vrije Universiteit Amsterdam, Department of Medical Oncology, Cancer Center Amsterdam, Amsterdam, The Netherlands

[2] Fondazione Pisana Per La Scienza, Pisa, Italy

[3] Amsterdam UMC, Univ of Amsterdam, Laboratory for Experimental Oncology and Radiobiology, Amsterdam, The Netherlands

[4] Amsterdam UMC, Vrije Universiteit Amsterdam, Department of Surgery, Amsterdam, The Netherlands

[5] U.O. Anatomia ed Istologia Patologica II Azienda Ospedaliero Universitaria Pisana , Pisa, Italy

[6] Oncode Institute, Amsterdam, The Netherlands

1 Introduction

Pancreatic ductal adenocarcinoma (PDAC) is the most common tumor type of the pancreas with a five-year survival rate not exceeding 8% [1]. A lack of reliable markers for early diagnosis, as well as its aggressive metastatic spread are the main causes of this extremely poor survival rate [2, 3]. The development of next-generation sequencing (NGS) has enabled detailed analyses of genomic aberrations and dysregulated gene expression patterns that underlie tumor development and progression, with KRAS, TP53, CDKN2A and SMAD4 as major oncogenic drivers of this disease. As yet, a comprehensive proteomic analysis of clinical PDAC samples is missing.

In recent years, multiple statistical methods and freely available bioinformatics tools have been developed that can extrapolate important features from high-throughput data, e.g.

pinpointing genes associated with clinical parameters such as cancer status or patient survival [4]. In this context, networks based on co-expression data [5] have extensively been used to identify densely interconnected genes associated with phenotypic traits. Most of the available algorithms have been applied to microarray- and RNAseq-based expression data [6, 7]. Using these approaches Tang et al. [8], for example, identified new prognostic markers in breast cancer. Additionally, these approaches have been used to search for potential therapeutic targets in small-cell lung carcinoma [9]. More recently, an integrative analysis of co-expression networks from proteomics and transcriptomics data in Alzheimer's disease revealed protein-specific networks in both asymptomatic and symptomatic patients [10, 11].

Weighted gene co-expression network analysis (WGCNA) assumes that the strength of node-to-node connections is best quantified by measures derived from their correlations. In co-expression networks for biological data, we refer to nodes as "genes" or "proteins". A glossary of network-related terms is reported in Table 1. Constructing co-expression networks is an effective way to characterize correlation patterns among nodes and to infer new biological functions of densely interconnected nodes called "modules". Modules can be related to external sample traits such as patient survival, recurrence and disease/health state, in order to discover biomarkers or therapeutic targets. Such modules are indicated by the Module Eigenprotein (ME; with a size of 1×20 in the current cohort, this is the most representative vector of values for that module) that can be related to external sample traits. In summary, the goals of a WGCNA analysis are: (i) establishment of real associations between proteins (instead of associations based on previous findings), (ii) identification of pathways specific for the dataset under analysis, (iii) association of modules to external information that provide biologically meaningful modules and (iv) identification of key drivers in relevant modules that may serve as candidate biomarkers and/or therapeutic targets.

Cancer proteomics aims to uncover the molecular basis of this devastating disease and to elucidate associated pathway features that cannot be detected by transcriptomics analyses. In this study, we report a PDAC proteomics analysis based on mass spectrometry (MS) data coupled to WGCNA to define networks of highly correlated proteins with specific functions associated with patient prognosis. We show that one module strongly features metabolic pathways and mesenchymal (EMT) signatures. This co-expression module was found to be significantly associated with disease-free survival (DFS) and overall survival (OS). The prognostic value of three key proteins in this module was validated in an independent cohort of 82 patients. These three proteins individually or in combination were able to predict patient survival.

Table 1 Glossary of network-related terms

Term	Definition	Reference
scale-free topology	Description of a cellular network structure in a graph theory concept	[20]
co-expression network	The edges are determined by the pairwise correlations between two protein expression profiles.	[6]
module	Module is a cluster of highly interconnected proteins.	[6]
connectivity	In co-expression networks, the connectivity measures how correlated a protein is with all other network proteins.	[6]
static tree cut	The branches of the hierachical clustering are cut at the same height. This is the most simple procedure for module identification.	[21]
dynamic tree cut	The module are defined by a non-constant cut off on the hierarchical clustering branches. This approach starts from a static tree cut and iteratively combine or remove proteins from one module to the other one. The iteration stops only when the modules reach stability.	[21]
adjacency matrix	Matrix containing pairwise correlations raised to the power β of all proteins.	[6]
weighted co-expression network	The edges of a network are described by weights. In the study, the weight is the correlation between two proteins raisd to the power β. This is essential to enhance strong correlations and avoid random noise.	[6]
unweighted co-expression network	Network that solely inform you if two proteins are connected or not	[6]
signed co-expression network	The edges of the network provide the sign of correlation (positive or negative)	[6]
unsigned co-expression network	The edges of the network do not provide the sign of correlation	[6]
direct network	The edges of this network described the action of one protein to another one (e.g. protein A is a kinase that phosphorylates protein B). It gives the direction of the action.	[6]
undirect network	The direction of the action is unknown.	[6]
module eigenprotein	The module eigenprotein ME is a vector with the most representative values of the given module and corresponds to the first principal component of that module.	[6]
hub gene	This term is used as an abbreviation of "highly connected gene" or specifically in this study, highly connected protein.	[6]

Explanatory table of the main network-related terms including references

2 Material and methods

2.1 Patient samples

Approval from the Local Medical Ethical committee at the VU University Medical Center was received (#14038). All patients provided informed consent for tissue sampling, clinical data analysis, and molecular analysis. Snap-frozen tumor samples from 39 patients included between January 2014 until November 2015 were evaluated by the Department of Pathology (Amsterdam UMC, Amsterdam). After pathological revision, 20 samples were eligible for further analysis. A minimum of 5–10% tumor surface was needed for further processing in this study. Clinical parameters were collected prospectively, OS and DFS data were obtained from electronic patient records. One patient was censored for OS analysis, since this patient succumbed to complications after surgery, defined as mortality within 60 days after surgery.

2.2 Protein isolation from bulk tumor tissue and sample preparation for mass spectrometry

Protein isolation was performed as previously described [12]. Briefly, protein lysates were separated on pre-cast 4%–12% gradient gels using the NuPAGE SDS-PAGE system (Invitrogen, Carlsbad, CA, USA). Gels were fixed in 50% ethanol/3% phosphoric acid solution, stained with Coomassie brilliant blue G-250 and then washed and dehydrated in 50 mM ammonium bicarbonate (ABC) once and 50 mM ABC/50% acetonitrile (ACN) twice. Gel lanes were cut into five bands, with each band sliced further into approximately 1 mm^3 cubes. The gel cubes were washed and dehydrated once in 50 mM ABC and twice in 50 mM ABC/50% ACN. Subsequently, the gel cubes were reduced in 10 mM DTT/50 mM ABC at 56 °C for 1 h, after which the supernatants were removed and the gel cubes were alkylated in 50 mM iodoacetamide/50 mM ABC for 45 min at room temperature in the dark. Next, the gel cubes were washed with 50 mM ABC/50% ACN, dried in a vacuum centrifuge at 50 °C for 10 min and covered with trypsin solution (Promega, 6.25 ng/ml in 50 mM ABC). Following rehydration with trypsin solution and removal of excess trypsin, the gel cubes were covered with 50 mM ABC and incubated overnight at 25 °C. Peptides were extracted from the gel cubes with 1% formic acid (FA) (once) and 5% FA/50% ACN (twice). All extracts were pooled and stored at −20 °C until use. Prior to liquid chromatography-mass spectrometry (LC-MS), the extracts were concentrated in a vacuum centrifuge at 60 °C, after which volumes were adjusted to 50 μl with 0.05% FA and filtered through a 0.45 μm spin filter into LC autosampler vials [13].

2.3 NanoLC-MS/MS proteomic analysis and database searching

NanoLC-MS/MS analysis was performed as previously described [14]. In brief, peptides were separated using an Ultimate 3000 nanoLC system (Dionex LC-Packings, Amsterdam, The Netherlands) equipped with a 40 cm × 75 μm internal diameter (ID) fused silica column custom packed with 1.9 μm 120 Å ReproSil Pur C18 aqua (Dr Maisch GMBH, Ammerbuch-Entringen, Germany). The samples were injected by gel band, starting with gel band 1 at the top of the gel for all samples, followed by gel band 2, until the final gel band 5. The experiment was considered as one continuous injection series with a blank injection at the start of the experiment. After injection, peptides were trapped at 6 μl/min on a 1 cm × 100 μm ID trap column packed with 5 μm 120 Å ReproSil C18 aqua at 2% buffer B (buffer A: 0.05% formic acid in MQ; buffer B: 80% acetonitrile +0.05% formic acid in MQ) and separated at 300 nl/min in a 10–40% buffer B gradient for 75 min (100 min inject-to-inject). Eluting peptides were ionized at a potential of +2 kVa into a Q Exactive mass spectrometer (Thermo Fisher, Bremen, Germany). Intact masses were measured at resolution 70.000 (at m/z 200) in the Orbitrap using an AGC target value of 3E6 charges. The top 10 peptide signals (charge-states 2+ and higher) were submitted to MS/MS in the HCD (higher-energy collision) cell (1.6 amu isolation width, 25% normalized collision energy). MS/MS spectra were acquired at resolution 17.500 (at m/z 200) in the orbitrap using an AGC target value of 1E6 charges, a maxIT of 60 ms and an underfill ratio of 0.1%. Dynamic exclusion was applied with a repeat count of 1 and an exclusion time of 30 s. MS/MS spectra were searched against a Swissprot reference proteome FASTA file (release January 2018, 42,258 entries, canonical and isoforms, no fragments), using MaxQuant version 1.6.0.16 [15]. Enzyme specificity was set to trypsin and up to two missed cleavages were allowed. Cysteine carboxamidomethylation (Cys, +57.021464 Da) was treated as fixed modification and methionine oxidation (Met, +15.994915 Da) and N-terminal acetylation (N-terminal, +42.010565 Da) as variable modifications. Peptide precursor ions were searched with a maximum mass deviation of 4.5 ppm and fragment ions with a maximum mass deviation of 20 ppm. Peptide and protein identifications were filtered at a false discovery rate (FDR) of 1% using the decoy database strategy. Proteins that could not be differentiated based on MS/MS spectra alone were grouped to protein groups (default MaxQuant settings). A protein was considered identified when at least 1 unique peptide was identified in one sample at high confidence (peptide and protein FDR < 1%). Searches were performed with the label-free quantification option selected. Proteins detected were quantified based on MaxQuant (version 1.6.0.16) output data. Label-free quantification (LFQ) intensities were filtered by contaminants and

only proteins with observations across all samples were retained. The MS proteomics data have been deposited to the ProteomeXchange Consortium via PRIDE [16] with accession number PXD015744.

2.4 Weighted gene correlation network analysis (WGCNA) and functional enrichment of identified modules

A protein co-expression network is an undirected graph, where each node corresponds to a protein, and each edge connects a pair of proteins that are significantly correlated [17]. The key concept in WGCNA is "connectivity". Connectivity describes direct and indirect relationships between two proteins/genes in networks [18].This metric has e.g. previously been used in breast cancer for drug prioritization by Neidlin et al. [19].

To investigate co-expressed proteins in resected patient PDAC samples, we used the WGCNA package [6] in R version 3.5.0 WGCNA defines modules as a group of densely interconnected proteins [18]. In unweighted networks the only information given is the correlation "yes" or "no", while for weighted networks, users can also gain information about the strength of a correlation. To remove random noise and enhance the strength of correlation, a particular threshold is required from the user. In the WGCNA package and so in this study, the choice was made by applying the scale-free topology criterion [20] using a soft threshold also known as "beta power". Different soft thresholds were tested (from 1 to 20) and power = 10 was retained to be enough to get an adjacency matrix very similar to a scale-free topology (correlation = 0.90) as shown in Supplementary Fig. S1. More explicitly, the adjacency matrix is obtained by using the correlation value between two proteins raised to the power threshold β (Formula 1)

$$a_{(ij)} = s_{(ij)}^{\beta}$$

where $a_{(ij)}$ is the weighted value of a protein in the adjacency matrix defined by rising the co-expression similarity $s_{(ij)}$ to a power β. Finally, modules are obtained by setting a cut-tree cut off on hierarchical clustering branches. In this study, a dynamic cut-tree method was chosen. Briefly, the algorithm starts by obtaining few large clusters by the static tree cut and, next, implements an iterative process of cluster decomposition and combination. The iteration stops only when the number of clusters becomes stable [21]. The dynamic tree method is essential to avoid relatively small modules. In this study, the minimum module size was set to 20 and the cut height was set to 0.998 automatically by the WGCNA tool. Each module was summarized by a vector of values (1 × 20 in this analysis, where 20 is the number of samples) that was called Module Eigenprotein (ME) and corresponded to the first principal component of the given module [22]. The final network was defined using the weighted option and threshold = 0.02 based on the value range of the data.

The GSEA of the modules was performed by mapping the proteins to gene names and submitting the gene list of each module to the GSEA Broad Institute browser [23]. Gene sets were ranked by significant p value and number of overrepresented genes. We adopted GSEA to perform functional enrichment analysis for each subnetwork based on GO biological process (BP) terms, cellular components (CC), hallmarks of cancer (HC) and transcription factor binding sites (TFBS). Each gene set was ranked using the FDR score and the number of overlapping genes between the module and the gene set. Moreover, the top 5 over-represented terms of each module were subjected to STRING analysis, in order to find the best descriptive biology for each specific module.

2.5 Survival analysis and meta-analysis

Clinical data of patients undergoing resection were obtained from electronic patient records and referral hospitals. Survival data were obtained from government registration. The ME for each module was then correlated to DFS and OS. Assuming to have a trait T and a Module Eigenprotein (ME), correlation or univariate regression models can be used to measure the extent of their association. Modules with a high trait significance may underlie biological pathways associated with the sample trait. Meta-analysis was carried out by applying univariate Cox regression, multivariate Cox regression and log-rank tests on our proteomics dataset and on two different transcriptomics datasets (TCGA [24] and Moffitt et al. [25]). Prognostic marker candidates were ranked based on the number of significant p values obtained from the above-mentioned statistical tests. Kaplan-Meier curves were plotted using "survminer" package in R.

2.6 Immunohistochemical validation of prognostic markers in an independent cohort

Immunohistochemistry (IHC) of tissue microarrays (TMAs) was evaluated as previously described [26]. In brief, FFPE tissues from resected patients were selected and combined in TMAs, including four representative cores from 4 different tumor areas for each patient. IHC staining of KHSRP, SPTBN1 and PYGL was performed according to the manufacturer's protocols. Anti-KHSRP monoclonal antibody (1:200, anti-KHSRP rabbit ab150393 Abcam), anti-SPTBN1 monoclonal antibody (1:500, anti-SPTBN1 mouse MA3–062, Invitrogen) and anti-PYGL polyclonal antibody (1:150, anti-PYGL rabbit ab198268) were used. Visualization was obtained using a BenchMark Special Stain Automation system (Ventana Medical Systems, Export, USA). Protein staining was evaluated by a molecular pathologist, assessing the

amount of tumor and tissue loss, background and overall interpretability. Cytoplasmic immunostaining intensity was classified into four grades: 0 (absent), 1 (weak), 2 (moderate) and 3 (strong), for both STPBN1 and PYGL. To reduce the scoring complexity, samples were defined as "with high expression", when the staining score was >2 in at least 50% of the tumor cells. The nuclear immunostaining intensity of KHSRP was classified into two grades: 0 (absent) and 1 (present). All patients provided written informed consent for the storage and analysis of their tumor material and survival data, respectively. This study was approved by the Local Ethics Committee of the University of Pisa (Ethics approval #3909, July 3rd, 2013).

Univariate and multivariate analyses were performed using a Cox regression model. Proteins with HR (Hazard Ratio) < 1 were considered protective and those with HR > 1 were defined as non-protective. Meanwhile, proteins with p values < 0.05 were considered statistically significant. A risk score method was used to assess the association of the three prognostic markers with OS in a multivariate analysis. The risk score was evaluated by combining the TMA scores of prognostic proteins weighted by their regression coefficients from univariate Cox regression (Formula 2).

$$Risk\ score = \sum_{i=1}^{n} \mathrm{TMA}_{\mathrm{score}} * \beta_{i.}$$

where n is the number of prognostic proteins, $\mathrm{TMA}_{\mathrm{score}}$ is the score of TMA for protein i, and β_i the regression coefficient of protein i in the univariate Cox regression analysis.

Group comparisons were evaluated using the unpaired nonparametric Mann-Whitney U test or unpaired Student's t test. Fishers exact test was used for categorical analysis. Correlations with clinicopathological characteristics, including DFS and OS, were tested using Kaplan-Meier curves and the log-rank test, as described above.

2.7 Cell culture

Hs766t cells (ATCC, Manassas, USA) were grown in DMEM (Lonza, Verviers, Belgium) supplemented with 10% heat-inactivated fetal bovine serum and 1% penicillin-streptomycin (10,000 U/ml, Gibco, Gaithersburg, MD, USA). Cells were kept at 37 °C in an atmosphere of 5% CO_2 in 75 cm^2 tissue culture flasks (Greiner Bio-One GmbH, Frickenhausen, Germany) and, for all the experimental procedures, harvested using trypsin-EDTA (Sigma, Zwijndrecht, The Netherlands) in their exponentially growing phase. Cells were tested within the last 3 months by microscopic morphology check and growth curve analysis according to the Cell Line Verification Test Recommendations (ATCC-Technical Bulletin No. 8, 2008). Periodic assays were carried out to detect mycoplasma contamination, and the identity of the cells was confirmed by PCR profiling using short tandem repeats (STR).

2.8 Immunofluorescence assay

Immunofluorescence analysis was performed according to a previously established protocol [27]. Briefly, cells were seeded in a Chamber-Slides System (Lab-Tek, Thermo Fisher Scientific, Waltham, USA) at a density of 5000 cells/well and allowed to attach overnight. Next, co-expression of KHSRP and SPTBN1 was evaluated in Hs766t PDAC cells, stained simultaneously with an anti-KHSRP monoclonal antibody (1:400, anti-KHSRP rabbit ab150393 Abcam) followed by an Alexa Flour 535 anti-rabbit antibody (Red; 1:70), and an anti-SPTBN1 monoclonal antibody (1:100, anti-SPTBN1 mouse MA3–062, Invitrogen) followed by an Alexa Flour 488 anti-mouse antibody (Green; 1:70). Nuclear DNA was stained with 4′, 6-diamidino-2-phenylindole (DAPI). Images were captured using a Zeiss Laser Scanning Microscope, processed and merged using Axiovision 4.1 software (Zeiss Microimaging, Thornwood, USA). In vitro experiments were performed with a minimum of three biological replicates, evaluating at least 100 cells.

3 Results

3.1 PDAC tissue proteomics and co-expression analysis

To obtain proteome level insight into PDAC cells, we used in-depth proteomics based on label-free nanoLC-MS/MS of gel-fractionated proteins to generate proteomic profiles of a cohort of 20 patients. The clinical characteristics of the selected patients are listed in Supplementary Table S1. We ensured equal protein loading of the samples to obtain optimal results (Supplementary Fig. S2). The obtained dataset consisted of 5667 proteins (contaminants removed) encoded by 5494 genes. Unsupervised clustering using all proteins did not reveal any specific grouping of the samples (Supplementary Fig. S3). The proteome dataset was subsequently used to establish a PDAC protein network. To obtain robust co-expression networks, we restricted the analysis to 993 proteins identified in all samples (Supplementary Table S2). Subsequent co-expression analysis yielded 12 consensus modules (Fig. 1), that were subsequently analyzed by GSEA to characterize the associated biology. Each module was annotated with gene sets and clinically relevant information. A complete list of genes associated with the modules is presented in Supplementary Table S3.

The modules covered a wide range of biological terms, and the most frequently occurring terms were those

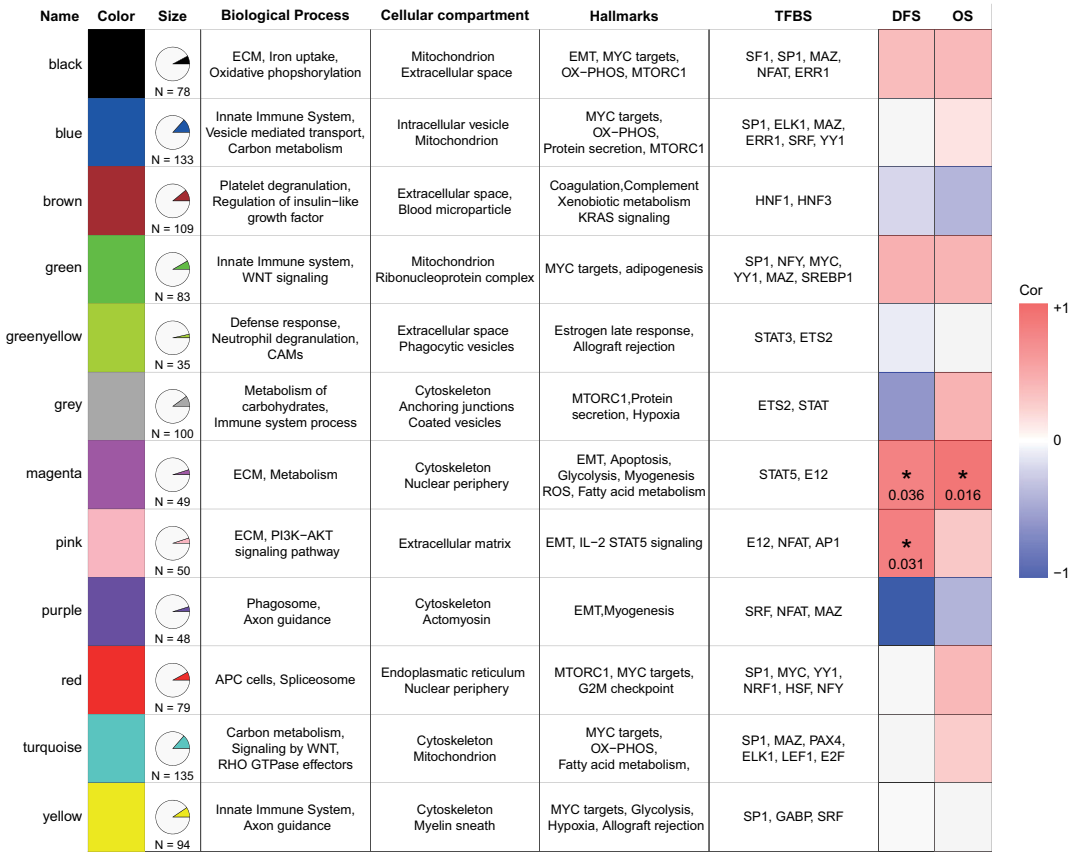


Fig. 1 Descriptive table of module characteristics. Module names, colors and numbers of proteins are indicated. Enrichment of biological processes, cellular compartments and hallmarks of cancer are described for each module by GSEA. The last two columns show correlations and significance levels of clinical endpoints for each module. The magenta module shows a positive (red color) and significant correlation with DFS (p value 0.036) and OS (p value 0.016) while the pink module shows a positive and significant correlation with DFS (p value 0.031)

implicated in metabolic processes in context of the mitochondrial compartment (black, green, magenta, turquoise and yellow modules). Furthermore, five modules (blue, green, green yellow, grey and yellow) consisted predominantly of immune system and defense response, probably regulated by STAT3 and ETS2 transcription factors (shown in the TFBS column in Fig. 1), while one module (brown) was linked to coagulation and platelet activation. Four modules were associated with epithelial-to-mesenchymal transition (EMT) processes (black, magenta, pink and purple modules). One module was enriched for transcription factors with STAT5A and E12 binding sites. However, these binding sites were predicted based on the binding regions present in the targets. The transcription factors did not show over-expression in our PDAC cohort.

3.2 Modules as potential prognostic markers for pancreatic cancer

The rationale behind the correlation network approach is to use the network language, which is particularly intuitive to biologists and allows for simple social network analogies. This method indeed allows the detection of biologically meaningful communities in the network and the study of relationships between them, helping the user to define interesting modules associated to external traits. Since co-expressed protein modules were identified and associated with hallmarks of cancer, we hypothesized that some modules may harbor potential markers for PDAC prognosis. Indeed, the magenta module, which presents EMT and glycolysis pathway components, exhibited positive and significant correlations with DFS and OS in our cohort (Fig. 1). Subsequently, we explored the

network biology of the prognostic co-expression module and found that this module is involved in metabolism, as can be inferred from the presence of PYGL, SOD2, GSR, GSS, PKM2, DDAH1 and TST, as well as EMT through ENO2, PLOD1 and FMOD (Fig. 2). Moreover, factors exclusively related to EMT were: COL12A1, TPM4, THBS1, FN1, POSTN, COMP, THBS2, CALU and FBLN2 (Fig. 2).

3.3 The Magenta module comprises candidate prognostic biomarkers for resected PDAC

For each module, we obtained the Module Eigenprotein (ME) for further survival association analyses. Only one module was significantly associated with DFS and OS in our proteomics cohort. The same protein signature was tested for OS association with transcriptomics data of the TCGA-PAAD project. The p value for all overall tests (i.e., Likelihood, Wald and Log Rank score) were 0.007, 0.01 and 0.01, respectively, indicating that the same gene signature is significantly associated to OS also in the transcriptomics data. Through subsequent investigation of epigenetic alterations of those genes, we found that oxidative stress and ECM-EMT related genes were not methylated. Therefore, we conclude that epigenetic inhibition of these genes was not prevalent. Additionally, we explored whether it was possible to refine the prognostic signature list by analyzing two publicly available independent transcriptomics datasets [24, 25]. To this end, different statistical tests were applied to prioritize the prognostic candidates, and the genes were ranked based on the frequency of significant observations among the tests (Supplementary Table S4). Our analysis revealed three potential top candidate biomarkers linked with prognosis: scaffold membrane protein spectrin beta chain, non-erythrocytic-1 (SPTBN1), splicing regulatory protein KHSRP and glycogen phosphorylase (PYGL) (Fig. 3). SPTBN1 is an actin-crosslinking protein that links the plasma membrane to the actin cytoskeleton. KHSRP is a multifunctional RNA-binding protein implicated in transcription, pre-mRNA splicing and mRNA localization to control important cellular processes such as metabolism, immune response, proliferation and differentiation. PYGL is a crucial phosphorylase that catalyzes the release of glucose molecules from glycogen, the major carbohydrate storage source. Cells under hypoxic conditions accelerate glycogen metabolism for an optimal glucose utilization (Warburg effect). Thus, PYGL is required for hypoxic cancer cells (as pancreatic cancer cells typically are) for glycolysis and glycogen degradation [28]. Based on genetic data from the TCGA consortium we found that alterations on PYGL can discriminate patient survival (p value < 0.001) even though the number of samples for short survival was relatively small. Interestingly, we found that high expression of SPTBN1 was associated with good prognosis in the proteomics data, but with poor prognosis in the transcriptomics data (Supplementary Fig. S4). Correlations between mRNA and protein data have been extensively studied and debated in the past years [29–31] and includes two recent large-scale clinical cancer proteogenomics studies. A more recent study by Vasaikar and colleagues [32] showed that enzymes belonging to the tricarboxylic acid (TCA) may be universally decreased at the protein level, but not at the mRNA level. This suggests a protein-level adaptation driving a strong Warburg effect in

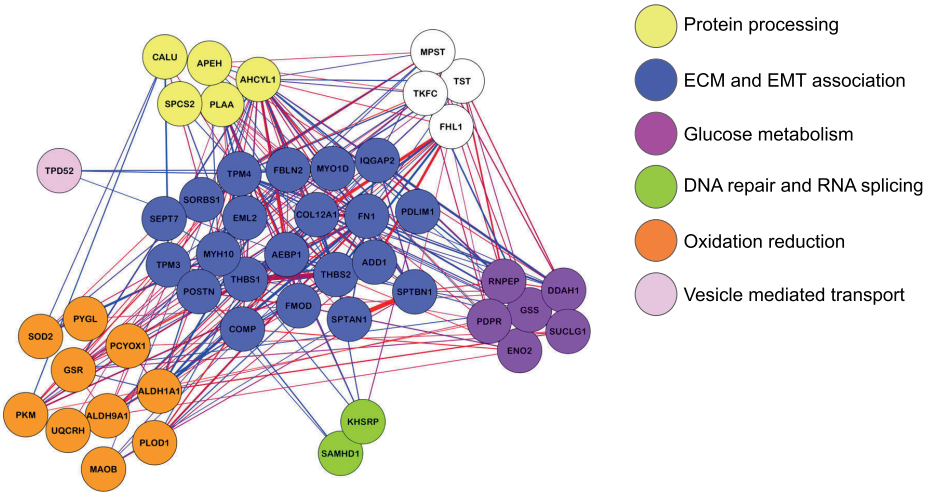


Fig. 2 Network visualization of the magenta module that associates with DFS and OS. Protein names are mapped to genes through Uniprot. Edge's widths represent the correlation strengths between genes (blue = negative correlation, red = positive correlation). Gene colors represent biological processes as indicated in the figure

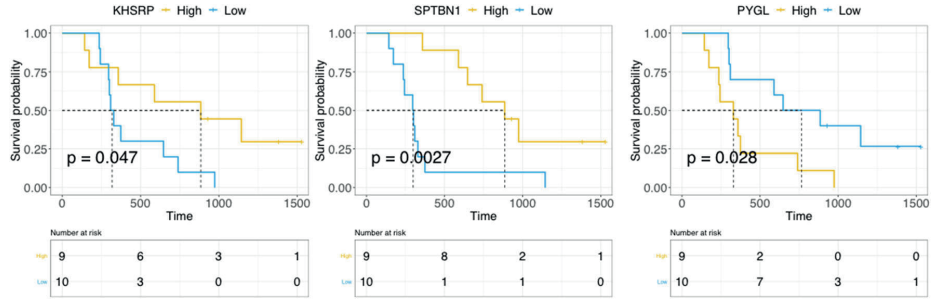


Fig. 3 Candidate biomarkers deduced from PDAC proteomics data. Kaplan-Meier curves of KHSRP, SPTBN1, KHSRP and PYGL proteomics data

microsatellite instable (MSI) colorectal cancer. In agreement with our study, the module where SPTBN1 belongs to is strongly enriched for metabolic genes. More specifically, these genes belong to glycolytic effects (PYGL, PKM, ENO2), thus preceding the TCA cycle. Another study by Mertins and colleagues [33] on breast cancer showed that despite a C-terminal truncation of GATA3, its protein expression level did not decrease, suggesting the occurrence of post-translational modification. Furthermore, these researchers found that signaling pathways such as PS1, ion channel transport and proteasome and basic cellular mechanism pathways, including ribosome, mRNA splicing, glycosylphosphatidylinositol biosynthesis and RNA polymerase, were enriched for negative correlations between mRNA and protein levels when compared to copy number alterations. Overall these findings suggest that post-translation modifications are more prone to occur in specific pathways compared to others. Since SPTBN1 has also been shown to carry genetic alterations in hepatocellular carcinoma patients with a short OS [34], this may be a starting point for future investigations on SPTBN1 mutations in PDAC patients.

3.4 Validation of KHSRP, SPTBN1 and PYGL as prognostic candidates for resected PDAC

We used WGCNA with unsigned networks. This means that the proteins in our modules can show both positive or negative correlations and that poor or good prognostic markers can fall in the same module because they are associated with the same biology. The top 3 prognostic markers, SPTBN1, KHSRP and PYGL, were chosen for subsequent IHC validation in an independent cohort of 82 resected PDAC patients (Fig. 4). Representative IHC images of tumor cores from two selected patients with highly divergent survival times and their SPTBN1, KHSRP and PYGL expression patterns are shown in Fig. 4a. In line with the proteome data, we found that SPTBN1 and KHSRP correlated with each other and were overexpressed in patients with good prognosis while PYGL, that anti-correlates with KHSRP and SPTBN1, correlated with poor prognosis. All three proteins had a significant prognostic value for OS (Fig. 4b) and PFS (Supplementary Fig. S5). Moreover, the signature of the three proteins taken together successfully predicted patient prognosis with $p = 0.0025$ (Supplementary Fig. S6). Co-expression of SPTBN1 and KHSRP was further confirmed by immunofluorescence in Hs766t cells. KHSRP (nuclear protein) and SPTBN1 (cytoplasmic protein) were clearly co-expressed (Fig. 4d) in the nucleus and in the cytoplasmatic compartment, respectively.

Finally, univariate and multivariate Cox regression models were used to assess the association of the three prognostic markers to OS. We found that SPTBN1, KHSRP and PYGL maintained significance in univariate and multivariate analyses when correcting for external factors. To assess whether all the three proteins were significantly associated with OS in a multivariate analysis, a risk score was evaluated showing that the prognostic signature of these three proteins was highly associated with OS in this independent cohort (Table 2).

4 Discussion

In the present study, we generated a comprehensive proteome dataset of 20 resected PDAC specimens and applied a weighted gene co-expression network analysis (WGCNA) to the data. WGCNA is a user-friendly and comprehensive software tool that has already been applied to several clinical features including brain cancer [35], diabetes [36] and chronic fatigue [37]. We focused on co-expression network analysis to infer biological functions and novel prognostic PDAC biomarkers. We reported a proteome dataset of 5667 proteins comprising 993 proteins identified in all samples giving rise to twelve modules in total. Protein co-expression modules were linked to well-known PDAC hallmarks of cancer such as axon-guidance, EMT, oxidative phosphorylation, MYC targets and KRAS signalling, as well as potential new relationships to biological processes. Importantly, one module was found to be significantly associated with survival. This module, called "magenta", was functionally enriched for glycolysis, EMT,

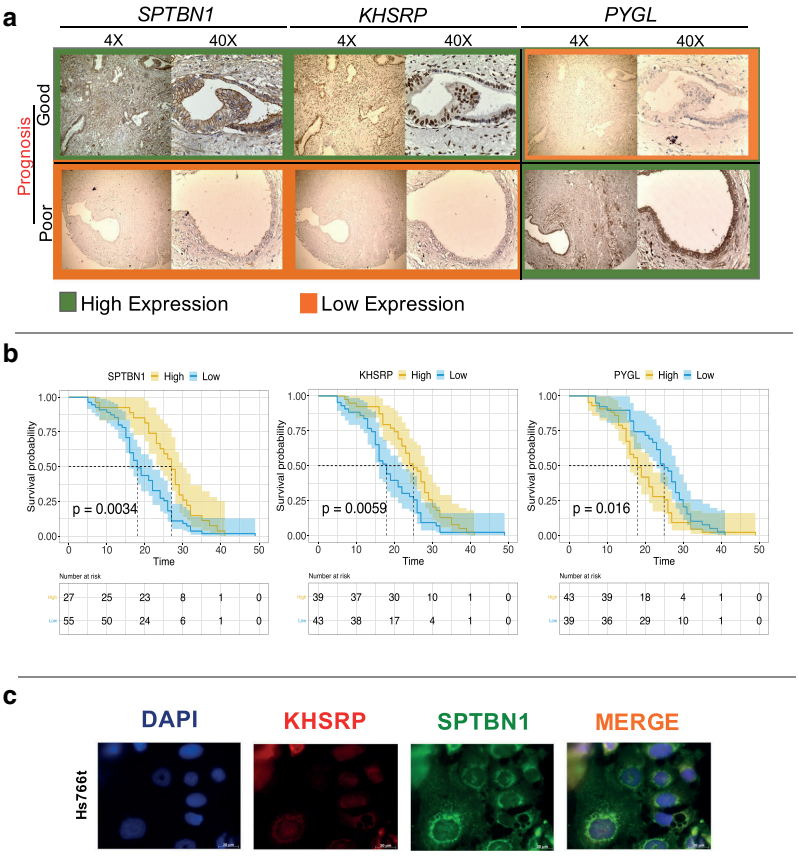


Fig. 4 SPTBN1, KHSRP and PYGL as prognostic markers for resected pancreatic cancer patients. **a**. Immunohistochemistry validation of SPTBN1, KHSRP and PYGL on TMAs of 82 patients. **b**. Kaplan-Meier curves for SPTBN1, KHSRP and PYGL with p values 0.0034, 0.0059 and 0.016, respectively. **c**. Immunofluorescence of KHSRP (red) and SPTBN1 (green) in Hs766t cells

apoptosis and reactive oxidative stress, highlighting a possible interplay between these biological processes.

Despite considerable experimental and computational modeling efforts, the role of EMT in cancer is still not fully understood [38]. In particular, the connection of EMT to various properties of cancer cells such as stemness, drug resistance, metabolism and metastasis is heavily discussed [39, 40]. In our current study, four modules showed overrepresentation of different sets of EMT genes (black, magenta, pink, purple), correlating with metabolic pathways, suggesting that cell metabolism can influence the EMT state or vice versa. Previously, tumor metabolism has been found to be associated with EMT [40, 41], illustrating the complexity of the interplay between EMT and metabolic reprogramming. Interestingly, these four modules were regulated by different transcription factors. In the magenta module transcription factors binding sites (TFBS) for STAT5 and E12 were noted. STAT5 has been shown to be overexpressed during EMT and aberrant activity of this transcription factor has been found to induce mitochondrial dysfunction and reactive oxygen species (ROS) formation, leading to DNA damage [42]. In addition, E12 has been found to be associated with repression of E-cadherin (and thus EMT) in mouse models [43].

Of note, all modules with SP1 as transcription factor binding site (yellow, turquoise, red, green, blue, black) where found to be associated with MYC targets as previously described [44, 45]. SP1 has been shown to regulate the expression of thousands of genes implicated in the control of a diverse array of cellular processes, such as growth [44], differentiation [46], apoptosis [44], angiogenesis [47] and immune response [48]. These cellular processes are all linked to the proteomic modules of our cohort that present SP1 as putative transcription factor.

The magenta module comprised three prognostic markers: SPTBN1, KHSRP and PYGL that were subsequently validated in an independent cohort of 82 patients. These markers may be used in the future to evaluate and

Table 2 Univariate and multivariate analysis of prognostic markers for resected PDAC

OS characteristics / protein expression		N	Median days (95% CI)	Univariate analysis	df	*P*-value		Multivariate analysis	df	*P*-value
Overall survival		82	22 (23.97 - 19.60)							
Age	<= 65	36	21.5 (25.87 - 19.62)	1.0 (ref)	1			NS*		
	> 65	46	22 (24.17 - 17.90)	1.13 (0.7 - 1.7)		0.57				
Gender	Female	44	21 (24.58 - 17.86)	1.0 (ref)	1			NS*		
	Male	38	23 (25.29 - 19.60)	0.97 (0.6 - 1.5)		0.9				
Grading	g1-g2	45	23 (26.95 - 20.78)	1.0 (ref)	1			1.0 (ref)	4	
	g3	37	21 (22.24 - 16.29)	1.73 (1.1 - 2.7)		**0.016 ***		1.44 (0.9 - 2.3)		0.125
LN status	no	12	25.5 (30.26 - 20.57)	1.0 (ref)	1			NS*		
	yes	70	21.5 (23.60 - 18.74)	1.36 (0.7 - 2.5)		0.32				
resection margin	no	50	22.5 (26.43 - 21.56)	1.0 (ref)	1			1.0 (ref)	4	
	yes	32	20 (22.25 - 14.43)	1.56 (0.9 - 2.4)		**0.052 .**		1.54 (0.9 - 2.4)		**0.075 .**
vascular infiltration	no	35	22 (27.83 - 21.36)	1.0 (ref)	1			1.0 (ref)	4	
	yes	47	22 (22.57 - 16.82)	1.56 (0.9 - 2.4)		**0.050 .**		1.26 (0.7 - 2.0)		0.331
SPTBN1	High	27	27 (22.23 - 17.14)	1.0 (ref)	1					
	Low	55	18 (29.79 - 22.35)	0.49 (0.3 - 0.7)		**0.003 ****				
KHSRP	High	39	25 (22.08 - 16.05)	1.0 (ref)	1		***Risk score***	1.0 (ref)	4	
	Low	43	18 (27.74 - 21.84)	0.53 (0.3 - 0.8)		**0.005 ****		1.47 (1.1 - 1.9)		**0.008 ****
PYGL	High	39	25 (27.41 - 21.19)	1.0 (ref)	1					
	Low	43	18 (22.49 - 16.52)	1.72 (1.1 - 2.6)		**0.016 ***				

Validation cohort characteristics with univariate and multivariate analyses for factors associated with OS. SPTBN1, KHSRP and PYGL remain significantly associated with OS together with grading stage, resection margin and vascular infiltration (significant *p* value in bold). In the multivariate analysis, significant covariates from univariate analysis are included and SPTBN1, KHSRP and PYGL are combined under the risk score

*NS: not significant in univariate cox regression; CI: Confidence of interval; df: degree of freedom

predict clinical responses of PDAC resected patients. SPTBN1 is a dynamic intracellular non-pleckstrin homology-domain protein, which plays important roles in cellular shape formation, protection of membranes against stress, positioning of transmembrane proteins, and molecular trafficking. Spectrin is made up of four subunits. Among these, the beta subunits are responsible for most of the binding activity and its role as a transforming growth factor-β signal transducing adapter protein that is necessary to form Smad3/Smad4 complexes [49]. KHSRP (KH-Type Splicing Regulatory Protein) controls important cellular processes such as proliferation, differentiation and metabolism. KHSRP (also known as FBP2) is a factor interacting with an enhancer element upstream of the *c-MYC* oncogene promoter [50]. In the past twenty years additional roles of KHSRP in post-transcriptional control of gene expression have been discovered with implications for pre-mRNA splicing [51], mRNA decay [52] and microRNA biogenesis [53]. PYGL catalyzes the degradation of glycogen [54] and is responsible for maintaining blood glucose homeostasis by regulating the release of glucose 1-phosphate from liver glycogen stores [55].

Importantly, transcriptomics data of all three biomarkers revealed significant associations with survival. Interestingly, SPTBN1 could be defined as a good prognostic marker based on the proteomics as well as the protein-based IHC data, while it was associated with poor prognosis based on the transcriptomics data (Supplementary Fig. S4A). Systematic studies have revealed multiple processes beyond the "non-correlation" of mRNA expression and protein concentration levels [56]. These include (i) specific translation rates of e.g. upstream open reading frames (uORFs) [57] or internal ribosome entry sites (IRES), (ii) translation rate modulation due to the binding of regulatory proteins or binding of micro-RNAs [58], (iii) modulation of a protein half-life involving the complex ubiquitin-proteasome pathway [59], or autophagy, which may influence protein concentrations independent of transcript levels.

Although we captured three new prognostic biomarkers and the biology associated with these, there are some limitations to our study that need to be noted. Due to the high heterogeneity of PDAC and the limited number of samples, we were not able to delineate proteomics-based PDAC subtypes. Exploring correspondence or correlation with known

PDAC subtypes is challenging due to the lack of PDAC subtypes based on proteomics data. Furthermore, because of the limited number of samples, this study should be considered as a first exploratory analysis and its prognostic relevance needs to be validated in additional clinical studies.

Taking together, our data indicate that an EMT-metabolic module is associated with the prognosis after surgical resection of PDAC patients and that the module's proteins SPTBN1, KHSRP and PYGL may serve as potential prognostic biomarkers. Our results also show that co-expression networks are able to extrapolate tumor-specific biology as well as biological mechanisms empowering prognostic marker discovery, even with a limited number of samples.

Acknowledgments The research reported in this publication was supported by the Dutch Cancer Society (KWF project #10212 and #11957, The Netherlands), an AIRC Start-Up grant (Italy), the CCA Foundation (Amsterdam, the Netherlands) and Fondazione Pisana Per La Scienza (Italy).

Author contributions GM performed bioinformatics analyses. AV supervised network biology mining. TLL performed sample preparations for LC-MS/MS. SP performed LC-MS/MS. TP acquired and processed mass spectrometry data. TLL and EG were involved in the selection of patient material and clinical data collection. JK, SP, CJ, TV and were responsible for experimental design and mass spectrometry. EG, MB and CRJ were involved in experimental design and manuscript preparation. NF and EG performed IHC validations. NF and MC performed immunofluorescence validations. CRJ and EG coordinated and supervised the study. All the authors critically reviewed the manuscript

Funding KWF grant Dutch Cancer Society (#10212 and #11957) (CJ, EG, MB), Italian Association for Cancer Research AIRC/Start-Up grant, Italy (EG), Fondazione Pisana Per La Scienza, Italy (EG).

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interests. MFB has received research funding from Celgene and has acted as a consultant for Servier. Neither were involved in the drafting of this manuscript, nor the design of this study.

Abbreviations WGCNA, weighted gene co-expression network analysis; PDAC, Pancreatic ductal adenocarcinoma; ME, Module Eigenprotein; EMT, Epithelial-mesenchymal-transition; OX-PHOS, Oxidative phosphorylation; TFBS, Transcription factor binding site; OS, Overall survival; DFS, Disease-free survival

References

1. P. Rawla, T. Sunkara, V. Gaduputi, Epidemiology of pancreatic cancer: Global trends, etiology and risk factors. World J. Oncol. **10**, 10–27 (2019)
2. A. Maitra, R.H. Hruban, Pancreatic cancer. Annu. Rev. Pathol. Mech. Dis. **3**, 157–188 (2008)
3. A. Vincent, J. Herman, R. Schulick, R.H. Hruban, M. Goggins, Pancreatic cancer. Lancet **378**, 607–620 (2011)
4. C. Wu, F. Zhou, J. Ren, X. Li, Y. Jiang, S. Ma, A selective review of multi-level Omics data integration using variable selection. High-Throughput **8**, 4 (2019)
5. A.-L. Barabási, R. Albert, Emergence of scaling in random networks. Science **286**, 509–512 (1999)
6. P. Langfelder, S. Horvath, WGCNA: An R package for weighted correlation network analysis. BMC Bioinformatics **9**, 559 (2008)
7. A.A. Margolin, I. Nemenman, K. Basso, C. Wiggins, G. Stolovitzky, R.D. Favera, A. Califano, ARACNE: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics **7** Suppl 1:S7 (2006)
8. J. Tang, D. Kong, Q. Cui, K. Wang, D. Zhang, Y. Gong, G. Wu, Prognostic genes of breast Cancer identified by gene co-expression network analysis. Front. Oncol. **8**, 374 (2018)
9. H. Nakamura, K. Fujii, V. Gupta, H. Hata, H. Koizumu, M. Hoshikawa, S. Naruki, Y. Miyata, I. Takahashi, T. Miyazawa, H. Sakai, K. Tsumoto, M. Takagi, H. Saji, T. Nishimura, Identification of key modules and hub genes for small-cell lung carcinoma and large-cell neuroendocrine lung carcinoma by weighted gene co-expression network analysis of clinical tissue-proteomes. PLoS One **14**, e0217105 (2019)
10. Q. Zhang, C. Ma, M. Gearing, P.G. Wang, L.-S. Chin, L. Li, Integrated proteomics and network analysis identifies protein hubs and network alterations in Alzheimer's disease. Acta Neuropathol. Commun. **6**, 19 (2018)
11. N. T. Seyfried, E. B. Dammer, V. Swarup, D. Nandakumar, D. M. Duong, L. Yin, Q. Deng, T. Nguyen, C. M. Hales, T. Wingo, J. Glass, M. Gearing, M. Thambisetty, J. C. Troncoso, D. H. Geschwind, J. J. Lah, A. I. Levey, A Multi-network approach identifies protein-specific co-expression in asymptomatic and symptomatic Alzheimer's disease. Cell. Syst. **4**, 60–72.e4 (2017)
12. F. Böttger, E.A. Semenova, J.-Y. Song, G. Ferone, J. van der Vliet, M. Cozijnsen, R. Bhaskaran, L. Bombardelli, S.R. Piersma, T.V. Pham, C.R. Jimenez, A. Berns, Tumor heterogeneity underlies differential cisplatin sensitivity in mouse models of small-cell lung cancer. Cell. Rep. **27**, 3345–3358.e4 (2019)
13. S.R. Piersma, J.C. Knol, I. de Reus, M. Labots, B.K. Sampadi, T.V. Pham, Y. Ishihama, H.M.W. Verheul, C.R. Jimenez, Feasibility of label-free phosphoproteomics and application to base-line signaling of colorectal cancer cell lines. Proteome Quest Understand Biol. Dis. HUPO **2014**, 247–258 (2015)
14. M. de Wit, H. Kant, S.R. Piersma, T.V. Pham, S. Mongera, M.P.A. van Berkel, E. Boven, F. Pontén, G.A. Meijer, C.R. Jimenez, R.J.A. Fijneman, Colorectal cancer candidate biomarkers identified by tissue secretome proteome profiling. J. Proteome **99**, 26–39 (2014)
15. J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. **26**, 1367 (2008)
16. J.A. Vizcaíno, E.W. Deutsch, R. Wang, A. Csordas, F. Reisinger, D. Ríos, J.A. Dianes, Z. Sun, T. Farrah, N. Bandeira, P.-A. Binz, I. Xenarios, M. Eisenacher, G. Mayer, L. Gatto, A. Campos, R.J. Chalkley, H.-J. Kraus, J.P. Albar, S. Martinez-Bartolomé, R. Apweiler, G.S. Omenn, L. Martens, A.R. Jones, H. Hermjakob, ProteomeXchange provides globally coordinated proteomics data submission and dissemination. Nat. Biotechnol. **32**, 223–226 (2014)

17. J.M. Stuart, A gene-Coexpression network for global discovery of conserved genetic modules. Science **302**, 249–255 (2003)
18. Z. Bin, H. Steve, A general framework for weighted gene co-expression network analysis. Stat. Appl. Genet. Mol. Biol. **4** Article17 (2005)
19. M. Neidlin, S. Dimitrakopoulou, L.G. Alexopoulos, Multi-tissue network analysis for drug prioritization in knee osteoarthritis. Sci. Rep. **9**, 15176 (2019)
20. R. Albert, Scale-free networks in cell biology. J. Cell Sci. **118**, 4947–4957 (2005)
21. P. Langfelder, B. Zhang, S. Horvath, Defining clusters from a hierarchical cluster tree: The dynamic tree cut package for R. Bioinformatics **24**, 719–720 (2008)
22. P. Langfelder, S. Horvath, Eigengene networks for studying the relationships between co-expression modules. BMC Syst. Biol. **1**, 54 (2007)
23. A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, A. Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, J.P. Mesirov, Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA **102**, 15545–15550 (2005)
24. R.L. Grossman, A.P. Heath, V. Ferretti, H.E. Varmus, D.R. Lowy, W.A. Kibbe, L.M. Staudt, Toward a shared vision for cancer genomic data. N. Engl. J. Med. **375**, 1109–1112 (2016)
25. R.A. Moffitt, R. Marayati, E.L. Flate, K.E. Volmar, S.G.H. Loeza, K.A. Hoadley, N.U. Rashid, L.A. Williams, S.C. Eaton, A.H. Chung, J.K. Smyla, J.M. Anderson, H.J. Kim, D.J. Bentrem, M.S. Talamonti, C.A. Iacobuzio-Donahue, M.A. Hollingsworth, J.J. Yeh, Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. Nat. Genet. **47**, 1168–1178 (2015)
26. E. Giovannetti, Q. Wang, A. Avan, N. Funel, T. Lagerweij, J.-H. Lee, V. Caretti, A. van der Velde, U. Boggi, Y. Wang, E. Vasile, G.J. Peters, T. Wurdinger, G. Giaccone, Role of CYB5A in pancreatic cancer prognosis and autophagy modulation. J. Natl. Cancer Inst. **106** (2014)
27. O. Firuzi, P.P. Che, B. El Hassouni, M. Buijs, S. Coppola, M. Löhr, N. Funel, R. Heuchel, I. Carnevale, T. Schmidt, G. Mantini, A. Avan, L. Saso, G.J. Peters, E. Giovannetti, Role of c-MET inhibitors in overcoming drug resistance in spheroid models of primary human pancreatic cancer and stellate cells. Cancers **11**, 638 (2019)
28. E. Favaro, K. Bensaad, M.G. Chong, D.A. Tennant, D.J.P. Ferguson, C. Snell, G. Steers, H. Turley, J.-L. Li, U.L. Günther, F.M. Buffa, A. McIntyre, A.L. Harris, Glucose utilization via glycogen phosphorylase sustains proliferation and prevents premature senescence in cancer cells. Cell. Metab. **16**, 751–764 (2012)
29. Y. Liu, A. Beyer, R. Aebersold, On the dependency of cellular protein levels on mRNA abundance. Cell **165**, 535–550 (2016)
30. M. Wilhelm, J. Schlegl, H. Hahne, A.M. Gholami, M. Lieberenz, M.M. Savitski, E. Ziegler, L. Butzmann, S. Gessulat, H. Marx, T. Mathieson, S. Lemeer, K. Schnatbaum, U. Reimer, H. Wenschuh, M. Mollenhauer, J. Slotta-Huspenina, J.-H. Boese, M. Bantscheff, A. Gerstmair, F. Faerber, B. Kuster, Mass-spectrometry-based draft of the human proteome. Nature **509**, 582–587 (2014)
31. A. Franks, E. Airoldi, N. Slavov, Post-transcriptional regulation across human tissues. PLoS Comput. Biol. **13**, e1005535 (2017)
32. S. Vasaikar, C. Huang, X. Wang, V.A. Petyuk, S.R. Savage, B. Wen, Y. Dou, Y. Zhang, Z. Shi, O.A. Arshad, M.A. Gritsenko, L.J. Zimmerman, J.E. McDermott, T.R. Clauss, R.J. Moore, R. Zhao, M.E. Monroe, Y.-T. Wang, M.C. Chambers, R.J.C. Slebos, K.S. Lau, Q. Mo, L. Ding, M. Ellis, M. Thiagarajan, C.R. Kinsinger, H. Rodriguez, R.D. Smith, K.D. Rodland, D.C. Liebler, T. Liu, B. Zhang, A. Pandey, A. Paulovich, A. Hoofnagle, D.R. Mani, D.W. Chan, D.F. Ransohoff, D. Fenyo, D.L. Tabb, D.A. Levine, E.S. Boja, E. Kuhn, F.M. White, G.A. Whiteley, H. Zhu, H. Zhang, I.-M. Shih, J. Bavarva, J. Whiteaker, K.A. Ketchum, K.R. Clauser, K. Ruggles, K. Elburn, L. Hannick, M. Watson, M. Oberti, M. Mesri, M.E. Sanders, M. Borucki, M.A. Gillette, M. Snyder, N.J. Edwards, N. Vatanian, P.A. Rudnick, P.B. McGarvey, P. Mertins, R.R. Townsend, R.R. Thangudu, R.C. Rivers, S.H. Payne, S.R. Davies, S. Cai, S.E. Stein, S.A. Carr, S.J. Skates, S. Madhavan, T. Hiltke, X. Chen, Y. Zhao, Y. Wang, Z. Zhang, Proteogenomic analysis of human colon cancer reveals new therapeutic opportunities. Cell **177**, 1035–1049.e19 (2019)
33. NCI CPTAC, P. Mertins, D.R. Mani, K.V. Ruggles, M.A. Gillette, K.R. Clauser, P. Wang, X. Wang, J.W. Qiao, S. Cao, F. Petralia, E. Kawaler, F. Mundt, K. Krug, Z. Tu, J.T. Lei, M.L. Gatza, M. Wilkerson, C.M. Perou, V. Yellapantula, K. Huang, C. Lin, M.D. McLellan, P. Yan, S.R. Davies, R.R. Townsend, S.J. Skates, J. Wang, B. Zhang, C.R. Kinsinger, M. Mesri, H. Rodriguez, L. Ding, A.G. Paulovich, D. Fenyö, M.J. Ellis, S.A. Carr, Proteogenomics connects somatic mutations to signalling in breast cancer. Nature **534**, 55–62 (2016)
34. J. Chen, S. Zaidi, S. Rao, J.-S. Chen, L. Phan, P. Farci, X. Su, K. Shetty, J. White, F. Zamboni, X. Wu, A. Rashid, N. Pattabiraman, R. Mazumder, A. Horvath, R.-C. Wu, S. Li, C. Xiao, C.-X. Deng, D.A. Wheeler, B. Mishra, R. Akbani, L. Mishra, Analysis of genomes and transcriptomes of hepatocellular carcinomas identifies mutations and gene expression changes in the transforming growth factor-β pathway. Gastroenterology **154**, 195–210 (2018)
35. S. Horvath, B. Zhang, M. Carlson, K.V. Lu, S. Zhu, R.M. Felciano, M.F. Laurance, W. Zhao, S. Qi, Z. Chen, Y. Lee, A.C. Scheck, L.M. Liau, H. Wu, D.H. Geschwind, P.G. Febbo, H.I. Kornblum, T.F. Cloughesy, S.F. Nelson, P.S. Mischel, Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. Proc. Natl. Acad. Sci. USA **103**, 17402–17407 (2006)
36. M.P. Keller, Y. Choi, P. Wang, D. Belt Davis, M.E. Rabaglia, A.T. Oler, D.S. Stapleton, C. Argmann, K.L. Schueler, S. Edwards, H.A. Steinberg, E. Chaibub Neto, R. Kleinhanz, S. Turner, M.K. Hellerstein, E.E. Schadt, B.S. Yandell, C. Kendziorski, A.D. Attie, A gene expression network model of type 2 diabetes links cell cycle regulation in islets with diabetes susceptibility. Genome Res. **18**, 706–716 (2008)
37. C. Priami, Algorithmic systems biology. Commun. ACM **52**, 80–88 (2009)
38. T. Brabletz, R. Kalluri, M.A. Nieto, R.A. Weinberg, EMT in cancer. Nat. Rev. Cancer **18**, 128–134 (2018)
39. K. Weidenfeld, D. Barkan, EMT and Stemness in tumor dormancy and outgrowth: Are they intertwined processes? Front. Oncol. **8**, 381 (2018)
40. C. Seliger, P. Leukel, S. Moeckel, B. Jachnik, C. Lottaz, M. Kreutz, A. Brawanski, M. Proescholdt, U. Bogdahn, A.-K. Bosserhoff, A. Vollmann-Zwerenz, P. Hau, Lactate-modulated induction of THBS-1 activates transforming growth factor (TGF)-beta2 and migration of glioma cells in vitro. PLoS One **8**, e78935–e78935 (2013)
41. G.V. Røsland, S.E. Dyrstad, D. Tusubira, R. Helwa, T.Z. Tan, M.L. Lotsberg, I.K.N. Pettersen, A. Berg, C. Kindt, F. Hoel, K. Jacobsen, A.J. Arason, A.S.T. Engelsen, H.J. Ditzel, P.E. Lønning, C. Krakstad, J.P. Thiery, J.B. Lorens, S. Knappskog, K.J. Tronstad, Epithelial to mesenchymal transition (EMT) is associated with attenuation of succinate dehydrogenase (SDH) in breast cancer through reduced expression of SDHC. Cancer Metab. **7**, 6 (2019)
42. C. Moser, P. Ruemmele, S. Gehmert, H. Schenk, M.P. Kreutz, M.E. Mycielska, C. Hackl, A. Kroemer, A.A. Schnitzbauer, O. Stoeltzing, H.J. Schlitt, E.K. Geissler, S.A. Lang, STAT5b as molecular target in pancreatic cancer–inhibition of tumor growth, angiogenesis, and metastases. Neoplasia N. Y. N. **14**, 915–925 (2012)
43. M.A. Pérez-Moreno, A. Locascio, I. Rodrigo, G. Dhondt, F. Portillo, M.A. Nieto, A. Cano, A new role for E12/E47 in the repression of *E-cadherin* expression and epithelial-mesenchymal transitions. J. Biol. Chem. **276**, 27424–27431 (2001)

44. J. Kaczynski, T. Cook, R. Urrutia, Sp1- and Krüppel-like transcription factors. Genome Biol. **4**, 206 (2003)
45. S. Kyo, M. Takakura, T. Taira, T. Kanaya, H. Itoh, M. Yutsudo, H. Ariga, M. Inoue, Sp1 cooperates with c-Myc to activate transcription of the human telomerase reverse transcriptase gene (hTERT). Nucleic Acids Res. **28**, 669–677 (2000)
46. O.G. Opitz, A.K. Rustgi, Interaction between Sp1 and cell cycle regulatory proteins is important in transactivation of a differentiation-related gene. Cancer Res. **60**, 2825 (2000)
47. N.M. Mazure, M.C. Brahimi-Horn, J. Pouysségur, Protein kinases and the hypoxia-inducible factor-1, two switches in angiogenesis. Curr. Pharm. Des. **9**, 531–541 (2003)
48. K. Jones, J. Kadonaga, P. Luciw, R. Tjian, Activation of the AIDS retrovirus promoter by the cellular transcription factor, Sp1. Science **232**, 755–759 (1986)
49. S. Chen, J. Li, P. Zhou, X. Zhi, SPTBN1 and cancer, which links? J. Cell. Physiol. **235**, 17–25 (2020)
50. T. Davis-Smyth, R.C. Duncan, T. Zheng, G. Michelotti, D. Levens, The far upstream element-binding proteins comprise an ancient family of single-strand DNA-binding Transactivators. J. Biol. Chem. **271**, 31679–31687 (1996)
51. H. Min, C.W. Turck, J.M. Nikolic, D.L. Black, A new regulatory protein, KSRP, mediates exon inclusion through an intronic splicing enhancer. Genes Dev. **11**, 1023–1036 (1997)
52. P. Briata, C.-Y. Chen, A. Ramos, R. Gherzi, Functional and molecular insights into KSRP function in mRNA decay. Biochim. Biophys. Acta **1829**, 689–694 (2013)
53. R. Gherzi, C. Chen, M. Trabucchi, A. Ramos, P. Briata, The role of KSRP in mRNA decay and microRNA precursor maturation. Wiley Interdiscip. Rev. RNA **1**, 230–239 (2010)
54. B. Burwinkel, H.D. Bakker, E. Herschkovitz, S.W. Moses, Y.S. Shin, M.W. Kilimann, Mutations in the liver glycogen phosphorylase gene (PYGL) underlying glycogenosis type VI (hers disease). Am. J. Hum. Genet. **62**, 785–791 (1998)
55. J.L. Ekstrom, T.A. Pauly, M.D. Carty, W.C. Soeller, J. Culp, D.E. Danley, D.J. Hoover, J.L. Treadway, E.M. Gibbs, R.J. Fletterick, Y.S.N. Day, D.G. Myszka, V.L. Rath, Structure-activity analysis of the purine binding site of human liver glycogen phosphorylase. Chem. Biol. **9**, 915–924 (2002)
56. C.J. McManus, G.E. May, P. Spealman, A. Shteyman, Ribosome profiling reveals post-transcriptional buffering of divergent gene expression in yeast. Genome Res. **24**, 422–430 (2014)
57. K. Wethmar, J.J. Smink, A. Leutz, Upstream open reading frames: Molecular switches in (patho)physiology. BioEssays **32**, 885–893 (2010)
58. L.W. Barrett, S. Fletcher, S.D. Wilton, Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. Cell. Mol. Life Sci. **69**, 3613–3634 (2012)
59. Y.-C. Tang, A. Amon, Gene copy-number alterations: A cost-benefit analysis. Cell **152**, 394–405 (2013)

Supplementary Appendix Chapter 3

Co-expression analysis of pancreatic cancer proteome reveals biology and prognostic biomarkers

Table S1. Clinicopathological data.

Feature	Grouping	N or value	Percentage (%)
	Total patients	20	
Age (years)	Mean	64.45	
	Range	50 to 78	
Sex	Female	9	45
	Male	11	55
Tumor differentiation	poor	9	45
	intermediate	9	45
	well	1	5
	unknown	1	5
Tumor percentage	< 50%	15	75
	>= 50%	5	25
Disease Stage	1A	1	5
	1B	1	5
	2A	3	15
	2B	8	40
	3	6	30
	4	1	5
The 8th T stage (AJCC)	T1	1	5
	T2	13	65
	T3	6	30
	T4	0	0
The 8th N stage (AJCC)	N0	5	25
	N1	9	45
	N2	6	30
The 8th M stage (AJCC)	M0	19	95
	M1	1	5
Adjuvant therapy	No adjuvant	9	45
	Gemcitabine	11	55
DFS (days)	Median	296	
	Range	105 to 595	
OS (days)	Median	366	
	Range	143 to 1530	

Table S2. Protein expression data
Table S3. Modules Genes

Table S4. Candide Biomarkers.

	PROTEIN DATA			RNA-SEQ DATA				
GENE	**uni-COX**	**LOG RANK**	**multi-COX**	**uni-COX TCGA**	**LOG RANK TCGA**	**TCGA SCAN**	**R2 Moffitt**	**Freq**
KHSRP	**0.010**	**0.047**	**0.03**			**0.01**	**0.029**	**5**
SPTBN1	**0.006**	**0.0027**	**6.43e-7**			**0.023**	**0.032**	**5**
PYGL		**0.028**		**0.003**	**0.003**	**0.00026**		**4**
COL12A1			5.39e-9	0.04		0.0012	0.0005	4
COMP		0.05	0.002			0.022	0.043	4
SEPT7			0.004		0.04	0.036	0.02	4
CALU				0.008		0.0027	0.0003	3
FN1	0.04					0.014	0.0003	3
TPM3			2.8e-5	0.004		0.00036		3
TPM4				0.006	0.029	0.00031		3
ALDH1A1	0.04		2.29e-14			0.045		3
MYO1D	0.024		8.14e-7				0.00003	3
AEBP1	0.016	0.045					0.043	3
FMOD	0.02	0.04				0.017		3
THBS2		0.05				0.04	0.0002	3
ADD1		0.016				0.00039		2
DDAH1						0.03	0.0004	2
PLAA	0.02					0.032		2
SPCS2				0.01		0.012		2
THBS1		0.05	2.00E-16					2
TST	0.016					0.04		2
PKM			0.01			0.000052		2
AHCYL1	0.003	0.009						2
MYH10			2.34e-10			0.0014		2
POSTN			0.0008			0.013		2
ALDH9A1						0.00046		1
PCYOX1		0.014						1
PLOD1						0.0042		1
RNPEP					0.048			1
SOD2						0.028		1
SPTAN1	0.004							1
SUCLG1						0.024		1
TPD52		0.0067						1
SAMHD1			0.00015					1
PDPR							0.049	1
FBLN2			3E-10					1
APEH								0
GSS								0

	PROTEIN DATA			RNA-SEQ DATA				
GENE	uni-COX	LOG RANK	multi-COX	uni-COX TCGA	LOG RANK TCGA	TCGA SCAN	R2 Moffitt	Freq
PDLIM1								0
EML2								0
ENO2								0
UQCRH								0
IQGAP2								0
MAOB								0
TKFC								0
GSR								0

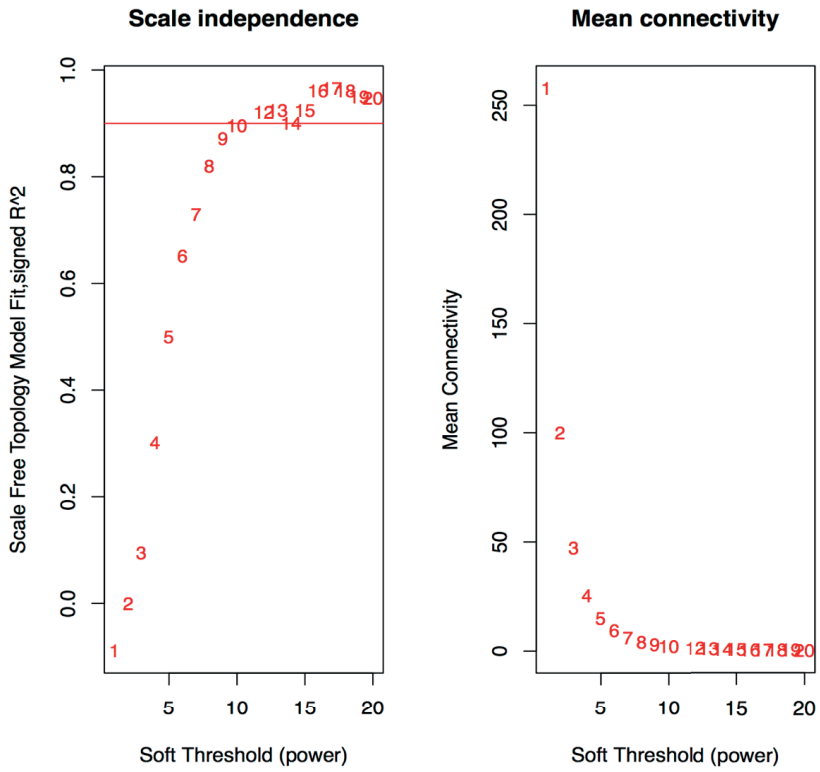


Figure S1. A sequence of numbers from 1 to 20 possible beta thresholds were chosen to test the correlation with the putative network from WGCNA and the scale-free topology network (plot on the left side). Description of the mean connectivity of the resulted network in relation to the beta-threshold chosen is shown in plot on the right.

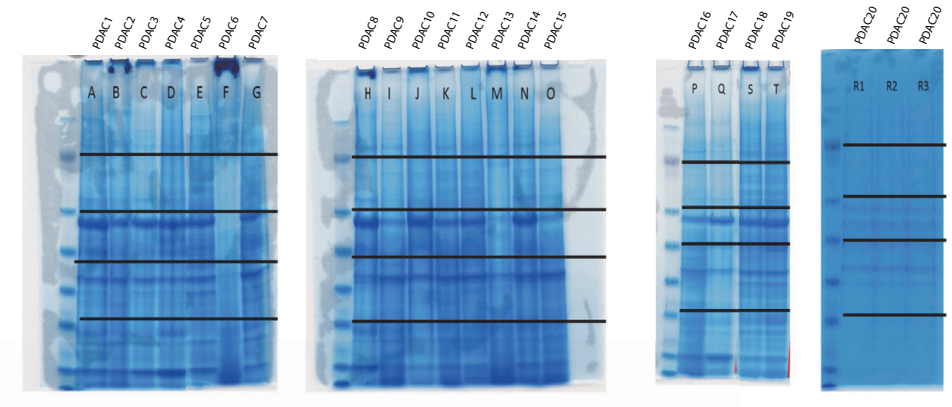


Figure S2. Coomassie staining of protein lysate to ensure equal protein load.

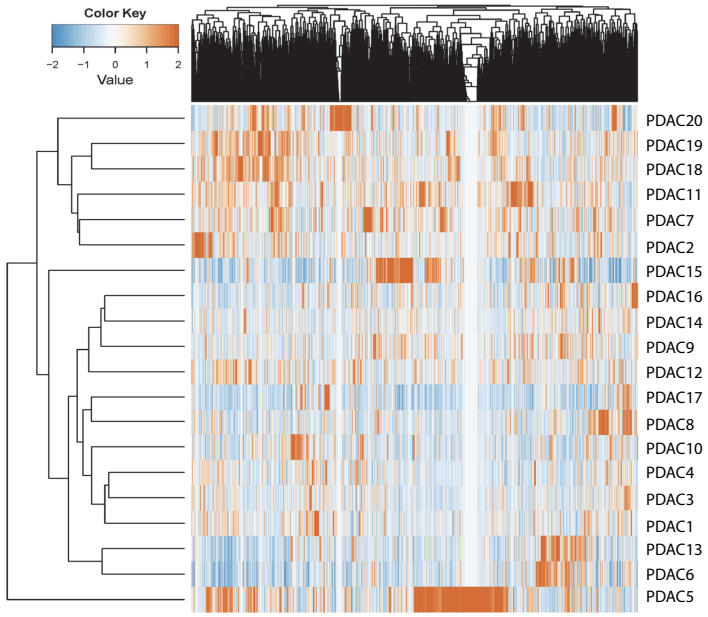


Figure S3. Unsupervised clustering of 20 resected pancreatic cancer patients using Euclidean distance and complete linkage. The unsupervised clustering was performed with the total number of proteins (N = 5764).

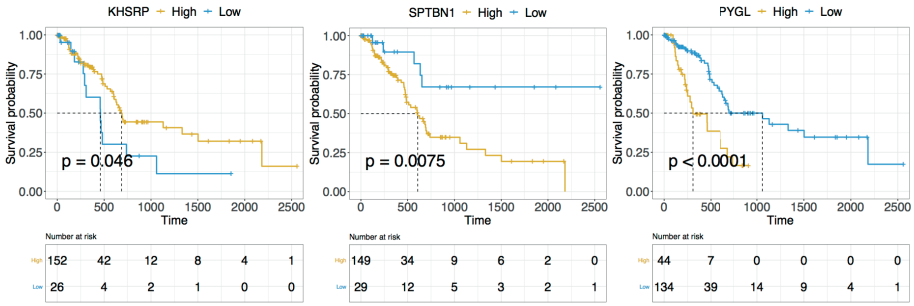


Figure S4. KM curves for KHSRP, SPTBN1 and PYGL on TCGA data

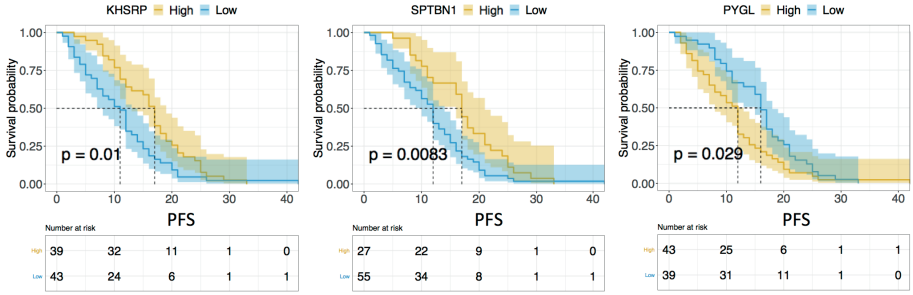


Figure S5. KM curves for KHSRP, SPTBN1 and PYGL on PFS data of validation cohort

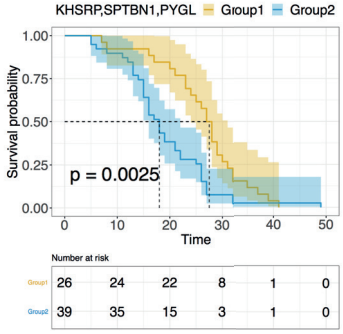


Figure S6. KM curve for KHSRP, SPTBN1 and PYGL taken together. Group 1 defines patients with high expression of SPTBN1 and KHSRP, and low expression of PYGL; Group 2 defines the opposite, low expression of SPTBN1 and KHSRP, and high expression of PYGL

Chapter 4

Microdissected pancreatic cancer proteomes reveal tumor heterogeneity and therapeutics

Le Large TYS, Mantini G, Meijer LL, Pham TV, Funel N, van Grieken NC, Kok B, Knol J, van Laarhoven HW, Piersma SR, Jimenéz CR, Kazemier G, Giovannetti E*, Bijlsma MF*
*These authors contributed equally

JCI Insights, 2020

JCI insight

RESEARCH ARTICLE

Microdissected pancreatic cancer proteomes reveal tumor heterogeneity and therapeutic targets

Tessa Y.S. Le Large,[1,2,3,4] Giulia Mantini,[2,4,5] Laura L. Meijer,[1,2] Thang V. Pham,[2,4] Niccola Funel,[6] Nicole C.T. van Grieken,[6] Bart Kok,[1] Jaco Knol,[2,4] Hanneke W.M. van Laarhoven,[7] Sander R. Piersma,[2,4] Connie R. Jimenez,[2,4] G. Kazemier,[1] Elisa Giovannetti,[2,5] and Maarten F. Bijlsma[3,8]

[1]Department of Surgery and [2]Department of Medical Oncology, Amsterdam University Medical Centers, Free University Amsterdam, Cancer Center Amsterdam, Amsterdam, Netherlands. [3]Laboratory for Experimental Oncology and Radiobiology, Amsterdam University Medical Centers, University of Amsterdam, Cancer Center Amsterdam, Amsterdam, Netherlands.[4]OncoProteomics Laboratory, Amsterdam University Medical Centers, Free University Amsterdam, Cancer Center Amsterdam, Amsterdam, Netherlands. [5]Cancer Pharmacology Lab, Fondazione Pisana per la Scienza, Pisa, Italy. [6]Unit of Anatomic Pathology II, Azienda Ospedaliera Universitaria Pisana, Pisa, Italy. [7]Department of Medical Oncology, Amsterdam University Medical Centers, University of Amsterdam, Cancer Center Amsterdam, Amsterdam, Netherlands. [8]Oncode Institute, Amsterdam, Netherlands.

Pancreatic ductal adenocarcinoma (PDAC) is characterized by a relative paucity of cancer cells that are surrounded by an abundance of nontumor cells and extracellular matrix, known as stroma. The interaction between stroma and cancer cells contributes to poor outcome, but how proteins from these individual compartments drive aggressive tumor behavior is not known. Here, we report the proteomic analysis of laser-capture microdissected (LCM) PDAC samples. We isolated stroma, tumor, and bulk samples from a cohort with long- and short-term survivors. Compartment-specific proteins were measured by mass spectrometry, yielding what we believe to be the largest PDAC proteome landscape to date. These analyses revealed that, in bulk analysis, tumor-derived proteins were typically masked and that LCM was required to reveal biology and prognostic markers. We validated tumor CALB2 and stromal COL11A1 expression as compartment-specific prognostic markers. We identified and functionally addressed the contributions of the tumor cell receptor EPHA2 to tumor cell viability and motility, underscoring the value of compartment-specific protein analysis in PDAC.

Authorship note: EG and MFB contributed equally to this work.

Conflict of interest: MFB has received research funding from Celgene and acted as a consultant to Servier. HWMVL has a consultant/advisory role with Bristol-Myers Squibb, Lilly, MSD, Nordic Pharma, Novartis, and Servier. HWMVL received research funding from Bayer, Bristol-Myers Squibb, Celgene, Janssen, Lilly, Nordic Pharma, Philips, Roche, and Servier.

Submitted: March 25, 2020
Accepted: June 24, 2020
Published: August 6, 2020.

Reference information: *JCI Insight*. 2020;5(15):e138290. https://doi.org/10.1172/jci.insight.138290.

Introduction

Large-scale omics efforts to identify key mediators of tumor biology have typically focused on the tumor compartment as an entity by itself, and tissue samples for gene expression and genomic analyses are commonly selected based on high tumor purity (1). This bias toward the epithelial compartment understates the influence of the microenvironment and its heterocellular composition (2). Further, by studying tumor tissue in bulk, the composition of specific compartments within the tumor can be masked. Bioinformatics efforts to delineate the compartment-specific expression have overcome this problem partially and increased our understanding of tumor biology (3). However, the inference of cellular composition by bioinformatics remains an estimation that requires validation, for instance, by the analysis of physically separated compartments. Especially in light of current molecular subtyping efforts to improve survival prediction and achieve a personalized treatment schedule, the characterization of the specific expression profiles of tumor cells and their microenvironment is key to identify patient groups more effectively and to discover therapeutic targets.

The relevance of the tumor-stroma interaction is of particular importance in tumors that are characterized by high stromal content, such as pancreatic ductal adenocarcinoma (PDAC). PDAC typically consists of small tumor islands surrounded by abundant extracellular matrix (ECM) and stromal cells, which together constitute the vast majority of the tumor bulk (4). This stroma is known to drive multiple protumorigenic features in PDAC tumor cells; it stimulates proliferation and metastatic growth, creates an immune-suppressive environment (5), and activates signaling via mechanobiology (6).

However, stromal depletion strategies had disappointing clinical results, suggesting that PDAC stroma also harbors tumor-restraining properties that suggest that a complex stromal plasticity exists (7–9).

PDAC responds poorly to systemic treatment, and only 9% of all patients diagnosed with PDAC survive more than 5 years (10). New chemotherapeutic regimens have improved the survival rate by several months (11, 12), and, hopefully, (neo)adjuvant multimodal approaches will improve survival further (13, 14). However, a better understanding of this tumor type is essential for the development of more effective, targeted therapeutic approaches and to select patients for such therapies. Several gene expression–based classification efforts have delineated the heterogeneity that is thought to contribute to the typical poor outcomes (15, 16), but most studies have largely ignored the separate stromal and epithelial contributions. Moffitt et al. (17) established molecular PDAC subtypes specific to the epithelial and stromal compartment by using an in silico deconvolution method on the transcriptomic landscape of bulk tumor tissue. Recently, Maurer et al. (18) performed in-depth transcriptomics of laser-capture microdissection (LCM) of PDAC, validating stromal and epithelial subtypes. These data provided evidence of the differential expression and pathways in both tumor cells and stroma and their divergent effects on prognosis. This compartment-specific complexity in PDAC warrants further understanding of each contribution to the disease outcome.

To date, no large-scale proteomic studies have been performed to identify the tumor- and stroma-specific protein landscapes of PDAC. Importantly, LCM can select compartment-specific areas and highly enrich the analysis for specific cell types (19). So far, LCM has predominantly been coupled with genomic or transcriptomic analysis, whereas proteomic studies of LCM material are rare due to the laborious efforts required to yield enough protein from the samples, especially from the stromal compartment (20). With the recent advances in sensitivity, throughput, and scan rate of liquid chromatography–mass spectrometry–based (LC-MS–based) proteomics (21), coupling LCM with LC-MS has become an attractive approach for identifying the functional proteomic landscape of PDAC and unraveling compartment-specific biology.

We hypothesized that tumor and stromal proteomic landscapes can yield complementary molecular information downstream of the transcriptome. Furthermore, we postulated that these data can be leveraged to dissect compartment-specific biology with prognostic value and to reveal targetable pathways. In this study, we report an extensive data set from LCM PDAC tissue and patient-derived xenografts (PDXs) and identify over 6000 proteins.

Results

Clinicopathological evaluation and depth of analysis. Of the 39 evaluated primary tumor samples, 16 were selected for LCM after confirmation of PDAC histology, with selection based on tissue quantity, tissue quality, and a tumor percentage of at least 5%. Additionally, bulk tumor analysis was performed on 11 samples matched with the LCM samples as well as on 5 unmatched bulk tumor samples. Baseline clinical characteristics of the total group (n = 21) are shown in Table 1 (additional information is provided in Supplemental Tables 1 and 2; supplemental material available online with this article; https://doi.org/10.1172/jci.insight.138290DS1). Three patients showed no sign of recurrence at last follow-up visit. The median overall survival (OS) of the cohort was 358 days. Age, sex, disease stage, and nodal involvement showed no statistically significant association with OS in this surgical cohort, likely due to small group sizes. Eleven patients received adjuvant systemic therapy, which associated with improved OS (log-rank test, P = 0.04). Tumor purities ranged from 5% to 70% (Supplemental Figure 1A), reflecting the known abundance of, and variance in, stromal content in PDAC.

A visualization of the workflow is shown in Figure 1A. Large areas of tumor and adjacent stromal areas were successfully isolated by LCM (Figure 1B). Single-shot LC-MS/MS analysis identified a total of 6214 unique proteins, yielding the deepest proteome data set of PDAC to date to our knowledge. Technical replicates were highly correlated (Supplemental Figure 1B), confirming reproducibility of the workflow. We identified 1866 unique epithelial proteins and 220 unique stromal proteins (Supplemental Figure 1C), indicating that protein expression is most heterogeneous in the tumor cell compartment.

Principal component analysis (Figure 1C) confirmed successful separation of the compartments, and this was supported by unsupervised clustering (Figure 1D) and correlation analysis (Supplemental Figure 1D). Interestingly, bulk tumors were more similar to the stromal samples than to the tumor compartment, regardless of tumor purity. Pearson correlation analysis confirmed the significantly higher correlation of bulk samples to compared with that of bulk samples to the tumor compartment (P < 0.0001, Supplemental Figure 1E). This result is in agreement with the high fraction of stroma in these tumors and emphasizes the caveats

JCI insight

Table 1. Baseline characteristics of the cohort

Clinicopathological characteristics	(*n* = 21)
Age at time of diagnosis, yr, median (range)	69 (50–80)
Sex, no. (%)	
Female	11 (52.4)
Male	10 (47.6)
CA19.9 (U/mL), median (range)	319.5 (5–12,639)
ASA classification, no. (%)	
1	1 (4.8)
2	14 (66.7)
3	6 (28.6)
Treatment, no. (%)	
Resection of the primary tumor	21 (100)
Disease stage (AJCC 8th edition)	
1B, no. (%)	1 (4.8)
2A, no. (%)	3 (14.3)
2B, no. (%)	9 (42.9)
3, no. (%)	7 (33.3)
4, no. (%)	1 (4.8)
Adjuvant therapy, no. (%)	
No adjuvant therapy	10 (47.6)
Gemcitabine	11 (52.4)
Disease-free survival, d, median (95% CI)	
Overall survival, d, median (95% CI)	369 (2–736)

of bulk tumor analyses. Moreover, interpatient correlation of tumor compartment samples was significantly lower (mean $R^2 = 0.5731$) than interpatient correlation of stromal samples (mean $R^2 = 0.7377$; $P < 0.0001$, Supplemental Figure 1F), indicating that the tumor compartment is the predominant source of heterogeneity in PDAC. This can partly be explained by the lower protein variety of the stromal compartment, as is shown from the lower number of different proteins expressed in stroma (Supplemental Figure 1C), but we propose that this is caused primarily by the relative genetic instability in the tumor compartment.

Biological characterization of epithelial and stromal proteomes. Unsupervised clustering using known markers for epithelial and stromal cells (Figure 2A) confirmed successful separation of compartments. This was also evident from per-sample analysis of selected tumor marker proteins (Supplemental Figure 1G). To further validate our LCM approach and compartment specificity, proteome data from 10 previously established PDAC PDXs (22) were generated. In these models, mouse cells and matrix proteins rapidly replace the human stroma, and the limited homology in amino acid sequence between mouse host and human donor can be leveraged to identify stromal and tumor cell–derived proteins. Species-specific unique proteins were allocated to a human or mouse protein set if they were identified in 5 of 10 samples. This approach yielded 1159 human- and 355 mouse-specific proteins, which we take to represent tumor and stroma proteins, respectively (Supplemental Data File 5). Single-sample gene set enrichment analysis (ssGSEA) showed strong correlation of mouse proteins to the stromal proteins from the LCM approach. Likewise, human proteins associated with the epithelial/tumor LCM compartment, but this enrichment was less evident (Figure 2B). The latter is likely caused by human PDX (tumor cell) proteins that are also expressed in stroma, diluting the relative enrichment. Additionally, ssGSEA of in silico identified stromal and tumor subtypes (17) showed an enrichment pattern confirming compartment specificity. Two tumor samples showed lower enrichment for both previously described Moffitt subtypes (17), implying the existence of tumor cell clones of other subtypes within these tissues or contamination by normal pancreatic tissue.

Regulation of protein expression downstream of RNA transcription is known to contribute to disparities between transcriptome and proteome data. To assess whether this is the case in models for PDAC, we correlated RNA-Seq data from the PDXs to their proteomics data and found an intermediate correlation of 0.55 between protein and mRNA transcripts, in line with previously reported data (Supplemental Figure 2A) (23). Of note, some proteins showed very high correlation to their transcripts, such as ENO1, while

JCI insight

RESEARCH ARTICLE

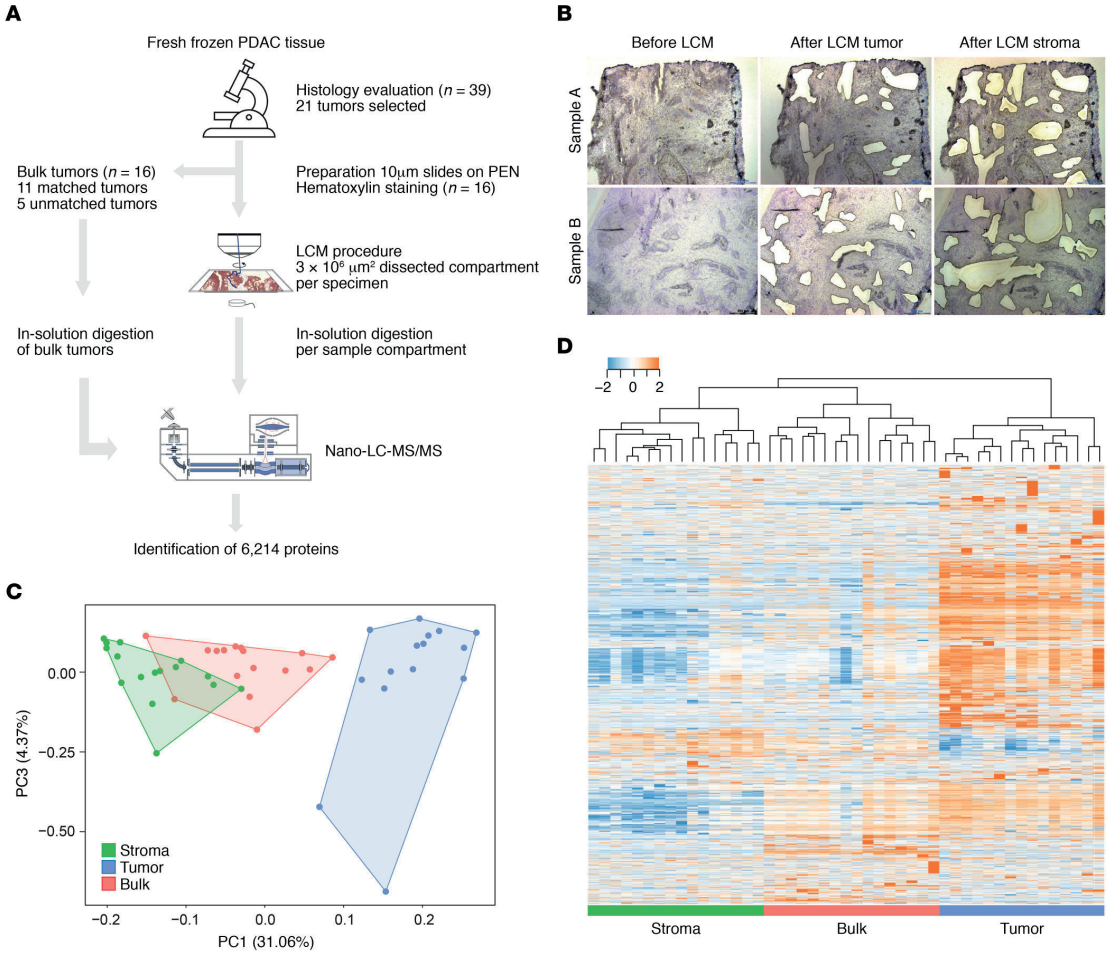


Figure 1. Proteome analysis of PDAC samples. (**A**) Flowchart of the experimental set up of the proteome landscape analysis (n = 16 for the LCM analysis, n = 16 bulk tumor analysis). (**B**) Representative example images of 2 PDAC tumors during the LCM procedure. Slides show before and after tumor microdissection and after stroma microdissection. Original magnification, ×10. (**C**) Principle component analysis (PCA) of MS/MS data shows separation of the epithelial compartment from stroma (n = 16), tumor (n = 15), and bulk samples (n = 11). (**D**) Unsupervised hierarchical clustering of all proteins shows compartment-specific expression of all samples (n = 42). Heatmap colors indicate relative expression levels.

others showed very poor correlation (PSMB1; Supplemental Figure 2, B and C). These findings suggest that a sizeable fraction of proteins are not accurately reflected in transcriptome data and underscore the added value of proteome analysis.

Next, we performed group GSEA of canonical pathways and biological processes. These analyses confirmed the divergent biological functions of tumor and the tumor microenvironment. Global pathway and signature analysis revealed characteristic pathways, such as ECM organization, complement cascade, and epithelial-mesenchymal transition, for the stroma (Figure 2C; Hallmark gene sets). In epithelial samples, enrichment of spliceosome and glycan biosynthesis signatures were found (Figure 2D; KEGG pathway gene sets). These associations confirm the biology known to underlie the epithelial and stromal compartments in PDAC. Moreover, the high enrichment for cancer cell populations also allowed the identification of less-explored biological processes. This includes, for example, oxidative phosphorylation, which was not recognized in bulk human tissue samples that largely reflect stromal GSEA signatures (Figure 2, C and D).

JCI insight

RESEARCH ARTICLE

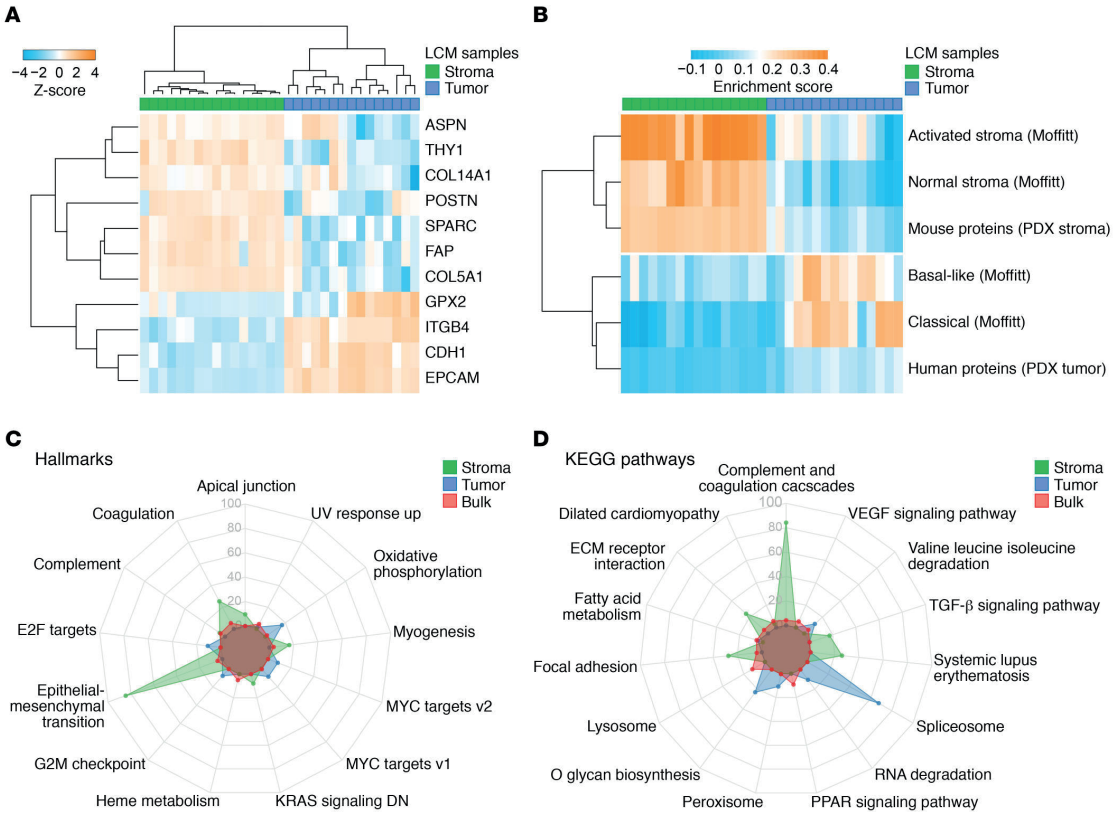


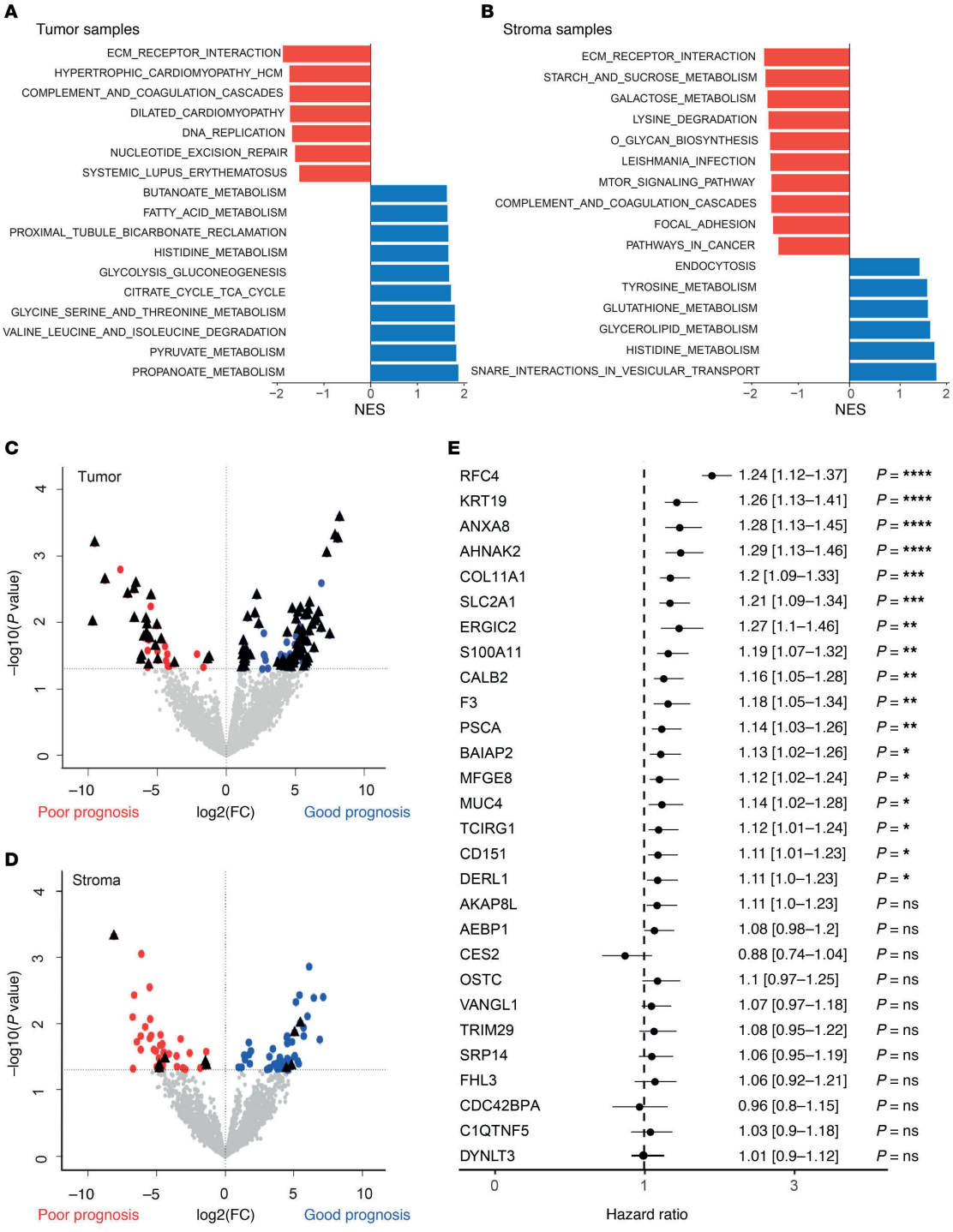
Figure 2. Successful LCM yields compartment-specific protein identification and biology enrichment. (**A**) Unsupervised hierarchical clustering of known tumor and stromal protein markers shows mutually exclusive expression patterns. (**B**) ssGSEA of previously published and newly created mouse/human PDX tumor and stroma gene sets (17) was performed, showing enrichment in concordance with compartment specificity. Data are shown in heatmap. Each column represents a tumor (n = 15) or stroma sample (n = 16). Each row represents a gene set that was used for enrichment. Colors indicate the enrichment score. (**C**) Radar plots of GSEA performed on "Hallmarks of cancer" and "KEGG pathways" (**D**) between sample origin reveal divergent biology of PDAC tumor and stroma, where bulk samples resemble stromal biology. Significant enriched gene sets were identified with GAGE (56), and the top 5 gene sets from each tissue origin with a P value of less than 0.05 are visualized.

Furthermore, unsupervised clustering on the most variable compartment-specific proteins showed 2 subclasses in the epithelial compartment, (Supplemental Figure 3A), of which cluster 2 showed a trend toward improved OS (Supplemental Figure 3B). Comparison of these groups with existing transcriptome tumor-specific subtypes revealed intermediate concordance (Supplemental Figure 3A), which might be due to sample size, precluding effective unsupervised clustering, or inapplicability of RNA-based classifications on enriched tumor compartment proteome data.

Additionally, a mixture of classical and basal-like protein expression in a single sample was identified in multiple tumors (Figure 2B), highlighting a degree of intratumor heterogeneity that likely results from the existence of populations of cells from different subtypes, as was recently described using gene expression data (24, 25). In the stroma compartment, 2 groups were identified (Supplemental Figure 3A), and these associated with the Moffitt et al. stromal subtypes on the proteome level (17). The normal stroma subgroup showed a trend toward better disease-free survival (DFS, log-rank test, P = 0.081, Supplemental Figure 3C). The activated stromal phenotype was identified more often than normal stroma, as described previously (n = 12 vs. n = 4, respectively) (17). In conclusion, these data reveal a profound heterogeneity in the proteome landscape of PDAC, with complex biological functionality of both tumor and stroma.

JCI insight

RESEARCH ARTICLE



JCI insight

Figure 3. Differential expression of short-term versus long-term survival reveals prognostic protein markers. GSEA of compartment-specific prognostic biology of (**A**) tumor and (**B**) stroma compartments. Red, poor prognosis gene sets; blue, good prognosis gene sets. The size of the bar indicates the significant normalized enrichment score (NES). Volcano plots shows differential proteins between poor (n = 6) and good (n = 6) prognosis patients in the tumor (**C**) and stromal compartment (**D**) ($P < 0.05$, limma test). Filtering on compartment specificity (cut-off at significance of $P < 0.1$, limma test). Black triangles indicate compartment-specific filtering. Colored dots represent significant proteins. (**E**) Forest plot of meta-analysis of compartment-specific prognostic markers from proteome analysis. Hazard ratio with confidence interval is visualized (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$, random effects model). Of the 28 proteins, 17 were validated as poor prognostic markers in transcriptomic data sets.

Compartment-specific proteins are associated with poor prognosis. Based on the results above, we hypothesized that LCM can enrich for prognostic proteins previously masked in bulk tissue analysis and that it can uncover tumor biology that underlies the poor disease outcome. To identify prognostic markers, patients with a short survival (n = 6) were compared with patients that survived more than 2 years following resection of the tumor (n = 6). These analyses showed divergent compartment-specific biology associated with disease outcome (Figure 3, A and B). Tumor areas from patients with a poor outcome were enriched for DNA replication and multiple cardiovascular signatures (Figure 3A), which represent proliferation and contractility. Importantly, tumor and stromal areas of patients with poor outcomes both showed enrichment for ECM receptor interaction, emphasizing the interaction of these compartments. Stroma from patients with a worse survival was enriched for focal adhesion pathway genes (Figure 3B), previously described as possible mediators of protumorigenic mechanosignaling (6). The majority of prognostic proteins identified in the stromal compartment by our approach were not specifically expressed in the microenvironment but were also found expressed in tumorous areas. After filtering for compartment specificity, 23 epithelial (Figure 3C) and 5 stromal proteins (Figure 3D) were found to associate with poor prognosis.

A meta-analysis based on previously published transcriptomic data sets was implemented for further validation of the identified markers (15–17, 26–28). Of 28 poor prognostic proteins identified, 17 were validated in the meta-analysis (Figure 3E), of which 15 were from the tumor and 2 from the stromal compartment, respectively. Additionally, of the 28 proteins, 16 were previously described in PDAC biology (Table 2), supporting the validity of our approach.

Next, prognostic stromal and epithelial markers were validated by IHC in 2 large independent cohorts of patients with resected PDAC (Supplemental Table 3). The epithelial marker calretinin/calbindin 2 (CALB2) was significantly correlated with OS in both cohorts (Figure 4, A and B; Supplemental Figure 4A; and Supplemental Table 3; $P = 0.009$ and $P = 0.006$, respectively). Predicted stromal marker collagen α-1(XI) chain (COL11A1) was significantly correlated with survival in the first independent cohort (Figure 4, C and D; $P = 0.016$); however, in the second validation cohort, no significance was observed (Supplemental Table 3 and Supplemental Figure 4B). This might be explained by the use of a tissue microarrays (TMAs) with cores selected for high tumor cell content that are therefore less representative of the stroma.

Identification of drug targets against PDAC. PDAC treatment currently depends on combinations of nontargeted cytotoxics (11, 12). The addition of tyrosine kinase inhibitors (TKIs) could expand the horizon for PDAC treatment, but currently available TKIs have not shown sufficient efficacy. We hypothesize that stromal proteins have masked targetable epithelial proteins in bulk proteome analyses, and we performed differential analysis of bulk samples versus the tumor compartment. Focusing on targetable receptor tyrosine kinases (RTKs), we identified 2 targetable RTKs in the LCM tumor cell compartment: epidermal growth factor receptor 1 (EGFR) and ephrin type-A receptor 2 (EPHA2; Figure 5A and Supplemental Figure 5A). EGFR has previously been targeted in an unselected population of patients with PDAC; the results of this study showed limited survival benefit (29). EPHA2 has not been targeted in clinical trials with patients with PDAC, whereas it has shown preclinical potential in modulating the immune evasive characteristic of PDAC (30). Indeed, a panel of PDAC cell lines showed variable EPHA2 expression (Supplemental Figure 5B), representative of the variable EPHA2 expression identified in our tumor compartment (Figure 5A). We selected 2 cell lines, Capan-2 and Hs766t, with high EPHA2 expression and phosphorylation to functionally evaluate EPHA2. Downregulation of EPHA2 by shRNAs (Supplemental Figure 5C) resulted in a reduced proliferation rate in Capan-2 cells (Figure 5B) but not in Hs766t cells (Supplemental Figure 5D), possibly explained by the higher relative phosphorylation of EPHA2 in the Capan-2 cell line (Figure 5D).

Recently, ALW-41-27, a EPHA2-specific inhibitor, was identified and was shown to target EPHA2 successfully in preclinical studies in non–small cell lung carcinoma (31). Exposure to this drug resulted in substantial inhibition of proliferation in a dose-dependent manner in PDAC cells (Figure 5C), with the strongest effect again found in Capan-2 cells, which have the highest phosphorylation of this kinase (Figure 5D).

Table 2. Compartment-specific prognostic proteins

Protein	FC	P value	Poor (*n*)	Good (*n*)	Meta-analysis	Known in PDAC	Staining on HPA
Stromal proteins							
AEBP1	1.4	0.0424	6	6	No	No	Both
C1QTNF5	8.1	0.0005	6	1	No	No	Stroma
COL11A1	4.8	0.0422	6	3	Yes	Yes (49)	NS
FHL3	4.8	0.0477	6	3	No	Yes (57)	NS
MFGE8	1.4	0.0440	6	6	Yes	No	Both
Tumor proteins							
AHNAK2	6.2	0.0369	6	3	Yes	Yes (58)	Tumor
AKAP8L	4.7	0.0182	3	0	No	No	Tumor
ANXA8	9.5	0.0006	6	1	Yes	Yes (59)	NS
BAIAP2	5.8	0.0109	6	2	Yes	No	Tumor
CALB2	6.6	0.0026	6	2	Yes	Yes (60)	Tumor
CD151	6.7	0.0032	4	0	Yes	Yes (61)	Tumor
CDC42BPA	6.0	0.0166	5	1	No	Yes (62)	Tumor
CES2	5.6	0.0436	5	1	No	Yes (63)	Tumor
DERL1	6.7	0.0087	4	1	Yes	Yes (64)	Tumor
DYNLT3	5.6	0.0175	3	0	No	No	Tumor
ERGIC2	5.0	0.0365	3	0	Yes	No	Tumor
F3	7.2	0.0038	4	0	Yes	Yes (65)	Tumor
KRT19	9.7	0.0098	5	1	Yes	Yes (66)	Tumor
MUC4	5.5	0.0040	6	1	Yes	Yes (67)	Tumor
OSTC	5.1	0.0228	6	2	No	No	NS
PSCA	8.8	0.0023	5	0	Yes	Yes (68)	ND
RFC4	5.8	0.0147	5	1	Yes	No	Tumor
S100A11	1.2	0.0336	6	3	Yes	Yes (69)	Tumor
SLC2A1	3.8	0.0412	4	1	Yes	Yes (70)	Tumor
SRP14	5.8	0.0089	6	2	No	No	NS
TCIRG1	1.4	0.0364	6	5	Yes	No	Tumor
TRIM29	6.1	0.0322	6	3	No	Yes (71)	Tumor
VANGL1	5.0	0.0111	4	0	No	No	Tumor

FC, $\log_2$-transformed fold change, patients with poor outcomes (Poor) vs. those with good outcomes (Good); HPA, Human Protein Atlas (http://www.proteinatlas.org/); ND, not detected; NS, no staining performed. Bold font indicates proteins after meta-analysis with significant associations (meta-analysis with random effects model).

Drug treatment resulted in a reduction of EPHA2 phosphorylation on the activating site phospho-Y588 within 2 hours (Figure 5D), confirming reduced activation. After 24 hours, surviving cells expressed less total EPHA2, indicating a possible selection for an EPHA2-negative cell population.

To determine the contributions of EPHA2 activity to tumor cell migration, Transwell migration assays were performed using EGF or FCS as chemoattractants. ALW-41-27 strongly inhibited migration (Figure 5E, $P = 0.0005$ and $P = 0.0029$, respectively). Interestingly, EPHA2-silenced cells were highly adherent to cell culture substrates and failed to release during trypsinization (Supplemental Figure 5E), providing further proof for a possible role in attachment and migration of these cells. Given the contributions of tumor cell migration to metastatic disease, these results and those from the proliferation assays suggest that the inhibition of EPHA2 could be very effective against PDAC in a clinical setting as well and underlines the usefulness of the current data set for exploration of PDAC biology.

Discussion

This study reports an in-depth proteomic analysis of LCM PDAC. It disentangles the stromal and tumor cell contributions to the proteome landscape of this aggressive tumor. The analyses performed show that the abundance of stroma in bulk tissue analysis masks important tumor biology and underscores the added value of compartment-specific analysis. We found that protein heterogeneity was

JCI insight

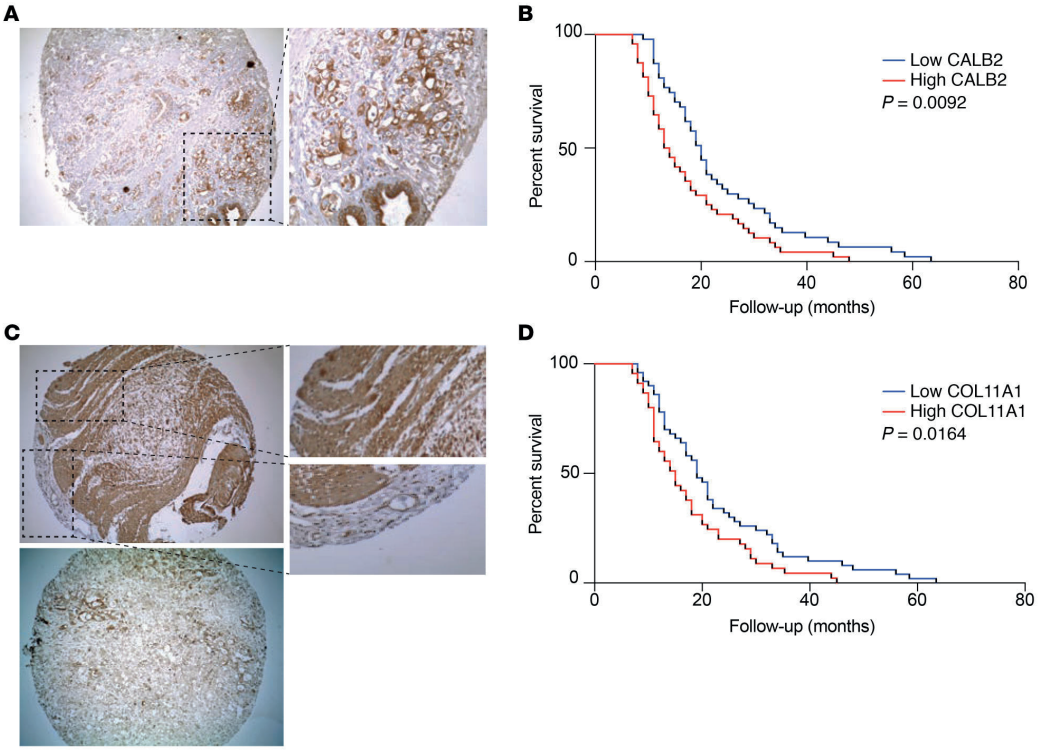
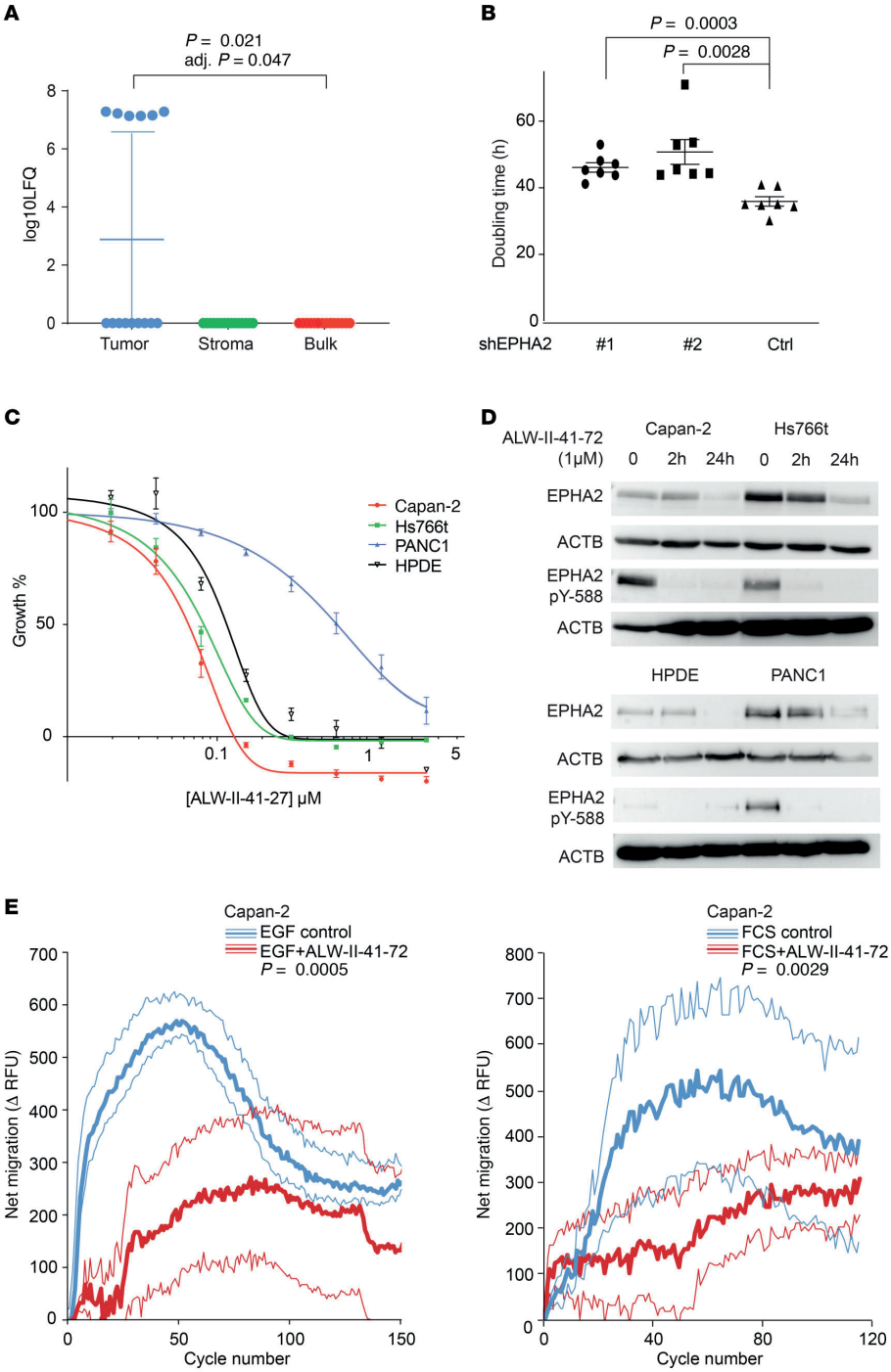


Figure 4. Validation of prognostic proteins CALB2 and COL11A1. (**A**) Representative IHC of TMA cores with CALB2-positive staining (original magnification, ×10 [left]; ×40 [right]). (**B**) Kaplan-Meier survival analysis of high (red, $n = 47$) or low (blue, $n = 48$) expression of CALB2 in the first independent cohort shows significant correlation to OS (median OS, 13.5 versus 20 months, $P < 0.0092$, log-rank test). (**C**) Representative IHC of TMA cores with CALB2-positive and -negative staining (original magnification, ×10 [left]; ×40 [right]). Additional views of a positive and negative area inside 1 tumor core are highlighted. (**D**) Kaplan-Meier analysis of high (red, $n = 50$) or low (blue, $n = 45$) expression of COL11A1 shows correlation in the first cohort (median OS, 15 versus 19 months, $P < 0.016$, log-rank test).

more pronounced in tumorous areas as compared with stroma and used this to identify prognostic markers and therapeutic targets.

PDAC is known to harbor relatively small tumor islands in vast areas of stroma, and the biology of both compartments effects multiple oncogenic pathways and thus patient outcome. Despite this knowledge, most efforts to unravel tumor biology have relied on genetic aberrations and transcriptomic networks (15–17, 27). Previous proteomic studies exploring PDAC used bulk tumor samples, performed LCM on very small sample sizes, did not reach the depth of proteome landscape currently described, or did not evaluate the stroma (32, 33). Using our cohort and pipeline, we were able to find a much higher degree of heterogeneity in the tumor compartment compared with the stroma. This can be explained by (a) the larger variety of unique proteins identified in tumor areas and (b) the fact that tumor cells are genetically unstable (compared with stromal cells) and that this is likely reflected in the heterogeneous proteome landscape (34).

Proteins known to associate with previously established RNA-based tumor subtypes (e.g., the Moffitt classification, ref. 17) were identified in our landscape. However, tumors were occasionally found to express proteins from both basal-like and classical subtypes. We propose that this results from the presence of mixed populations within a tumor, as was also found in recent single-cell RNA-Seq analyses of PDAC (25). Importantly, "normal" and "activated" stromal subtypes were fully recapitulated in our data, validating previously described stroma subtypes. Future studies on large cohorts are needed to validate the discriminative power of transcriptomic subtypes in proteome data.


A
P = 0.021
adj. P = 0.047
log10LFQ
Tumor
Stroma
Bulk
B
P = 0.0003
P = 0.0028
Doubling time (h)
shEPHA2
#1
#2
Ctrl
C
Capan-2
Hs766t
PANC1
HPDE
Growth %
[ALW-II-41-27] μM
D
ALW-II-41-72 (1μM)
Capan-2
Hs766t
0 2h 24h 0 2h 24h
EPHA2
ACTB
EPHA2 pY-588
ACTB
HPDE
PANC1
EPHA2
ACTB
EPHA2 pY-588
ACTB
E
Capan-2
EGF control
EGF+ALW-II-41-72
P = 0.0005
Net migration (Δ RFU)
Cycle number
Capan-2
FCS control
FCS+ALW-II-41-72
P = 0.0029
Net migration (Δ RFU)
Cycle number

JCI insight

Figure 5. EPHA2 phosphorylation in PDAC cells conveys sensitivity to the inhibitor ALW-II-27. (**A**) Differential expression of EPHA2 between bulk (n = 16) and isolated epithelial compartments (n = 15) identified EPHA2 as specifically but heterogeneously expressed in the tumor compartment (limma test, corrected for multiple testing, P = 0.05) (mean ± SD). (**B**) Reduction of EPHA2 expression by shRNA in Capan-2 cells significantly reduced proliferation rate (P < 0.001, unpaired 2-way t test) (mean ± SEM) (n = 7). (**C**) Dose-response curves of cell lines treated with ALW-II-27. (**D**) Western blot analysis of cells after treatment with ALW-II-27 shows reduction of EPHA2 expression and phosphorylation of activation site pY-588. See complete unedited blots in the supplemental material. (**E**) Transwell migration of Capan-2 cells was reduced upon pretreatment with 1 μM ALW-II-27, irrespectively of attractant (EGF, P = 0.0005, FCS P = 0.0029, 2-tailed t test with Welch's correction) (mean ± SEM) (n = 3).

Another proteome approach to evaluate extracellular matrix content is purification of the matrisome, a technique that measures enriched proteins after degradation of cellular content. However, it does not take into account the importance of cellular communication and warrants a large volume of bulk tumor samples, making it less appealing for biopsies or other samples of limited quantity. One study explored the matrisome of PDAC (35) and indicated that prognostic proteins originate mainly from the tumor cells, but this study did not measure contributions from both compartments. Importantly, 4 (AEBP1, C1QTNF5, COL11A1, and MFGE8) of the 5 potential prognostic stromal proteins identified in the current study were significantly upregulated in the PDAC matrisome compared with normal tissue, validating our results on proteome level. Large PDAC transcriptomic (RNA) data sets have been published, and we validated our candidate marker proteins in a meta-analysis of those gene expression data. We confirmed several known prognostic pancreatic markers and validated CALB2 as a poor prognostic tumor marker and COL11A1 as a poor prognostic stromal marker in PDAC. CALB2 has previously been described as a diagnostic marker and tumor-promoting protein in mesothelioma (36). It is associated with poor prognostic basal-like subtypes in other cancers (37). This protein plays an important role in intracellular calcium regulation, and it is necessary for vital cellular functions and cell contractility (38). Its role in cancer has not been fully explored; however, the poor prognostic value could relate to altered mechanobiology known to contribute to PDAC (6). The validated stromal prognostic marker demonstrated in this study, COL11A1, is a recently discovered effector of stromal activation and is suggested to play a pan-cancer role in tumor-stroma interaction (39).

Additionally, the comparison of bulk versus tumor-enriched samples showed the advantage of enrichment in proteomic studies, since these, as do genomic and transcriptomic studies, do not rely on amplification methods. For example, the presence of several kinases was masked in proteomes of bulk tumor. However, by enriching for tumor areas, we uncovered EPHA2 and functionally validated this kinase as therapeutic target in PDAC. shRNA knockdown was effective against cell growth in 1 cell line of 2 tested, which correlated with the phosphorylation of EPHA2. An additional explanation could be that, upon silencing, a bias is introduced for EPHA2-independent clones. A previously published study in which silencing of EPHA2 was achieved by siRNAs showed an antiproliferative effect on PDAC cells upon knockdown (40). Interestingly, EPHA2 was recently identified as a key regulator of the PDAC immune-suppressive microenvironment (30). Experimental work evaluating additional cell lines for the EPHA2-driven effect using inducible knockdown systems would strengthen study results; however, our study did revealed strong tumor-inhibiting and antimigration effects of the EPHA2 inhibitor ALW-41-27 in PDAC cells with high EPHA2 expression. This finding warrants further evaluation of this compound against PDAC.

There are several limitations to our study, the most important being the relatively small sample size. For instance, disease stage did not correlate significantly with survival due to small group size. However, this study remains the largest PDAC proteome landscape to date to our knowledge, and our approach added to the current knowledge on protein localization in PDAC and validated previously described transcriptomic stromal subtypes. In addition, despite the lack of association of outcome with staging, associations of survival with stromal proteome subtypes were strong. The limited sample size might also restrict evaluation of the full heterogeneity present in PDAC. This was apparent from the discordance between our unsupervised clustering analysis and the RNA-based tumor cell–specific classification.

A practical limitation was the surface needed to yield optimal proteome exploration. Adding to this is the fact that proteomics does not incorporate amplification steps, and low abundance proteins can be underrepresented. Therefore, to create the deepest data, identification and exclusion of possible singular tumor cells that may exist in the isolated stromal areas was not performed, which could contaminate the stromal protein landscape to some degree. Despite this limitation, we were able to create a PDAC data set of depth, with very little cross-compartment contamination.

The functional experiments were hampered by the limited number of cell lines tested, and future studies will have to be performed to formally prove the association of EPHA2 levels and phosphorylation with response to EPHA2 inhibitors. Despite this, our comprehensive exploration by IHC and functional investigation of markers/targets underscores the use of our data set for future studies.

In conclusion, we report a large-scale compartment-specific proteome landscape of PDAC. With this data set, future bulk tumor proteome data sets can be deconvolved and annotated. We have shown that our data can be the basis for in vitro analysis of targets of tumor signaling and bidirectional stimulatory communication between PDAC and stromal cells. Moreover, we have shown that prognostic genes can exist in tumor or its microenvironment or can be expressed by both. Our data underline the need to understand the biology of PDAC on multiple levels and warrant future large-scale proteome studies of PDAC.

Methods

Further details are provided in the Supplemental Methods.

Sample collection. Snap-frozen tumor samples were evaluated for quality and tumor percentage. The workflow is shown in Figure 1A. Sixteen samples were eligible for further LCM analysis and prepared as described previously (41). Additionally, unseparated bulk tumor of 11 matched and 5 unmatched samples was prepared for single-shot MS/MS analysis. Two patients suffered from postoperative complications and were excluded from survival analysis but included in tumor biology analysis.

LCM procedure. LCM was performed on the Leica LMD7000 instrument to yield a total surface density of 3×10^6 μm^2 cells per compartment. Selected areas were captured in 0.1% RapiGest SF Surfactant (Waters) and stored until further use. Samples were sonicated and reduced to a final concentration of 5 mM dithiothreitol (MilliporeSigma) and 15 mM iodoacetamide (MilliporeSigma). Sequencing-grade modified trypsin (Promega) was added overnight to a final concentration of 7 ng/μl. Digestion was stopped with acidification by trifluoroacetic acid (MilliporeSigma). The peptide mixture was centrifuged, and the supernatant was transferred to a glass-lined MS/MS auto sampler vials. Samples were brought to a volume of 20 μl and stored at –80°C until further analysis.

Peptide preparations of PDX models and bulk tumor tissue. Tissue from 10 previously established PDX models was used for proteome analysis (22). NSG mice were bred in-house. A minimum of 100 mg tissue was digested in lysis buffer (9 M urea, 20 mM HEPES pH 8.0, 1 mM Na_3VO_4, 2.5 mM $Na_4P_2O_7$, 1 mM $Na_2C_3H_7PO_6$). 50 μg protein was loaded on a NuPAGE 4%–12% gradient gel (Invitrogen). Proteins were digested and extracted according to our whole-in-gel protocol, as described previously (42). Bulk PDAC tumors were lysed, and in-solution digestion was performed according to the protocol of our laboratories after reduction and alkylation (43).

Nano-LC-MS/MS. Proteins were identified by nano-LC-MS/MS, as described previously (42). Peptides were separated by an Ultimate 3000 nano-LC-MS/MS system (Dionex LC-Packings). Eluting peptides were ionized into a Q Exactive HF mass spectrometer (Thermo Fisher Scientific). Intact masses were measured in the orbitrap using an AGC target value of 3×10^6 charges. The top 15 peptide signals were submitted to MS/MS in a higher-energy collision cell in MS/MS. MS/MS spectra were acquired at resolution 17.500 (at m/z 200) in the orbitrap.

Protein annotation and data analysis. MS/MS spectra were searched against the Swissprot FASTA file (LCM data: release March 2017; 42,161 entries, canonical and isoforms; proteome profiles from PdX data were searched with both human and mouse fasta files;uniprot_human_referenceproteome_2014_01_NO_fragments_42104entries.fasta, 61,552 entries); Uniprot_Mus_musculus_reference_proteome_2015_06_NO_FRAGMENTS_Canonical and isoforms _34331entries.fasta, 42,296 entries) using MaxQuant 1.5.8.0 (44). Enzyme specificity was set to trypsin, and up to 2 missed cleavages were allowed. Peptide and protein identifications were filtered at an FDR of 1% using the decoy database strategy. The minimal peptide length was 7 amino acids. A match-between-runs setting was implemented for analysis of low abundant proteins in the LCM database. For the PDX data, peptides mapped to solely human proteins were retained and subsequently collapsed in protein groups. Protein compartment specificity was correlated to mouse- and human-specific proteins identified from PDX PDAC tumors (protein had to be uniquely mouse or human specific). The mass spectrometry proteomics data have been deposited at the ProteomeXchange Consortium via the PRIDE (45) partner repository (data set identifiers PXD011289 and PXD017393). Protein and peptide data are available in Supplemental Data Files 1–8.

RNA-Seq of PDX models and correlation with protein expression. RNA-Seq of PDX was previously performed and described (EMBL-EBI ArrayExpress code E-MTAB-6830) (46). In short, total RNA was

JCI insight

extracted from PDX model tissue and amplified with the Total Prep RNA amplification kit (Illumina). RNA was sequenced on the Illumina HiSeq2500. Mapped reads were mapped to both mouse and human genomes. The RNA transcripts mapping solely the human genome were retained for further analysis. RPKM data were $\log_2$ transformed and filtered to include transcripts with an average read count ≥1 in all samples, obtaining 14,809 RNA transcripts in total. Human-specific spectral counts were $\log_2$ transformed and filtered to include proteins with an average spectral count ≥3 in all samples, obtaining 1705 protein groups in total. Spearman's correlation was applied to the resulting transcriptome-proteome matched file with 1627 genes (proteins were collapsed at gene level) to evaluate PDX transcriptome and proteome correlation. Methodology was described previously by Wang et al. (23).

In silico validation of prognostic markers. Data from publicly available transcriptomic (microarray or RNA-Seq) data sets (15–17, 26–28) with survival data were downloaded from GEObase, and each data set was scaled to a mean of 0, with a standard deviation of 1 to allow meta-analysis. Univariate cox proportional hazard regression models were evaluated for genes of interest. The Metafor R package (47) was used to perform meta-analysis validation of identified prognostic markers.

IHC validation on TMAs. Proteins of interest in both compartments, COL11A1 and CALB2, were evaluated in 2 independent cohorts (n = 95 and n = 95) by staining TMAs as described previously (48). Assessment of IHC staining of COL11A1 was performed in 3 tumor cores per patient containing representatives regions of the desmoplastic reaction. Expression was evaluated in relation to the stromal surface, as described previously (49). IHC of CALB2 was evaluated in the tumorous areas, taking into account positivity and intensity of the staining, as described previously (50).

Cell culture. PANC1 cells (ATCC) were cultured in RPMI medium (Lonza). Capan-2 cells (ATCC) and Hs766t cells (ATCC) were cultured in DMEM medium (Lonza). Media was supplemented with 10% heat-inactivated fetal calf serum (Biowest) and 1% penicillin-streptomycin (Lonza). An immortalized pancreatic ductal cell line (HPDE, supplied by Ming Tsao, Ontario Cancer Institute, Toronto, Ontario, Canada) (51) was cultured in supplemented KGM medium (Lonza).

Western blot validation of EPHA2 expression. Cell lysates from PANC1, Hs766t, Capan-2, and HPDE cells were created with diluted ×10 RIPA buffer (Abcam) containing protease and phosphatase inhibitors. 20 μg protein was used for Western blot. Primary antibodies were incubated in 5% BSA (MilliporeSigma) in PBST followed by secondary antibodies in 5% blocking buffer. Visualization was performed by an Uvitec Imaging station (Cleaver Scientific). A list of antibodies is provided in the Supplemental Methods.

Stable EPHA2 knockdown. Lentiviral plasmids were produced by transfecting HEK293T cells with EPHA2-targeting pLKO.1 constructs (MISSION shRNA Library clone numbers TRCN0000006403 and TRCN0000197131) and a scrambled control (shc002). Transfected supernatant was collected after 48 hours and filtered through a 0.45 μm filter (EMDMillipore). At 30% confluence, PDAC cell lines Capan-2 and Hs766t were transduced and subsequently selected after 48 hours with 2 μg/mL puromycin (MilliporeSigma). Knockdown efficiency was evaluated by Western blot as described above.

In vitro validation of drug target. Cells were plated in 96-wells plates and allowed to attach overnight. Growth was evaluated over 72 hours with respect to the control at the start of the experiment. To evaluate proliferation, the doubling time was calculated. A specific EPHA2 inhibitor, ALW-II-41-27 (APExBIO), was evaluated for cytotoxic effect. DMSO (MilliporeSigma) was used as control. Effect on cell proliferation was quantified with Sulforhodamine B (MilliporeSigma) protein staining as described previously (52). Migration was evaluated by Transwell migration as described before (53) following 15 minutes pretreatment. The AUC was calculated for each replicate of the migration curve.

Statistics. Data were analyzed with R (version 3.5.2.). Zeros were imputed based on normal distribution SD of log-transformed intensity data before differential expression analysis. Differential compartment expression was tested by paired statistics analysis (Limma package) (54) and Benjamini-Hochberg corrected for multiple testing. Two patients succumbed to postoperative complication (PDAC10 and PDAC17) and were excluded from the survival analysis. These samples were, however, included in the proteomics and downstream nonclinical analyses. Prognostic proteins were identified by group comparison (short OS, <1 year versus long OS, >2 years). GSEA (55) was performed in R. In vitro experimental comparisons were evaluated by paired or unpaired 2-tailed Student's t test. Correlation of clinicopathological characteristics and gene expression with DFS/OS were evaluated by Kaplan-Meier and log-rank test. The prognostic value of IHC scoring was tested with uni- and multivariate analysis. Error bars show the mean ± SEM. A P value of < 0.05 was considered statistically significant.

Study approval. The work described was carried out in accordance with the Declaration of Helsinki. Approval for tissue collection was obtained from the local medical ethical committees at the Amsterdam UMC (protocol 14038 for METC VUmc and 2011_126 and 201_181 for AMC, both Amsterdam, the Netherlands). Ethical approval for the validation cohorts was received from Comitato di Bioetica Azienza Ospedaliero-Universitaria Pisana, University Hospital of Cisanello, Pisa, Italy (protocol 37677; study no. 3909; July 3, 2013)). All patients provided informed consent for participation in the study. Animal work was performed in a previous study according to protocols approved by the animal experiment ethical committee at the Amsterdam UMC (protocol DTB102348, LEX102774).

Author contributions

TYSLL and NF performed LCM. TYSLL, GM, and TVP performed bioinformatic analysis. TYSLL, LLM, NF, and EG were involved with selection of patient material, clinical data collection, and statistical analyses. TYSLL, NF, and NCTVG performed pathological review of LCM material and performed IHC staining. JK, SRP, CRJ, and TVP were responsible for experimental design and mass spectrometry. TYSLL, HWMVL, GK, EG, and MFB were involved with experimental design and manuscript preparation. TYSLL, MFB, and BK performed experimental validation. CRJ, EG, and MFB coordinated and supervised the study. All the authors critically reviewed the manuscript.

Acknowledgments

We would like to acknowledge C. Maurer for his help with the meta-analysis of PDAC transcriptomic data, and all the patients participating in our studies that made this research possible. This work was supported by a Cancer Center Amsterdam Alliantie-AIO grant (to EG, MFB, and CRJ), the Bennink Foundation (to GK, EG, and TYSLL), the Dutch Cancer Society (10212 to CRJ, MFB, and EG), and an Italian Association for Cancer Research AIRC/Start-Up grant, Italy (to EG). The Netherlands Organization for Scientific Research (Middelgroot project no. 91116017, to CRJ) is acknowledged for supporting the mass spectrometry infrastructure.

Address correspondence to: Elisa Giovannetti, Amsterdam UMC, Vrije Universiteit Amsterdam, Department of Medical Oncology, De Boelelaan 1117, 1081 HV Amsterdam, Netherlands. Phone: 31.20.444.22.67; Email: e.giovannetti@amsterdamumc.nl. Or to: Maarten F. Bijlsma, Amsterdam UMC, University of Amsterdam, Laboratory for Experimental Oncology and Radiobiology, Meibergdreef 9, 1105 AZ Amsterdam, Netherlands. Phone: 31.20.5664824; Email: m.f.bijlsma@amsterdamumc.nl.

1. Yadav VK, De S. An assessment of computational methods for estimating purity and clonality using genomic data derived from heterogeneous tumor tissue samples. *Brief Bioinform*. 2015;16(2):232–241.
2. Pietras K, Östman A. Hallmarks of cancer: interactions with the tumor stroma. *Exp Cell Res*. 2010;316(8):1324–1331.
3. Yoshihara K, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun*. 2013;4:2612.
4. Wilson JS, Pirola RC, Apte MV. Stars and stripes in pancreatic cancer: role of stellate cells and stroma in cancer progression. *Front Physiol*. 2014;5:1–11.
5. Jiang H, et al. Targeting focal adhesion kinase renders pancreatic cancers responsive to checkpoint immunotherapy. *Nat Med*. 2016;22(8):851–860.
6. Laklai H, et al. Genotype tunes pancreatic ductal adenocarcinoma tissue tension to induce matricellular fibrosis and tumor progression. *Nat Med*. 2016;22(5):497–505.
7. Rhim AD, et al. Stromal elements act to restrain, rather than support, pancreatic ductal adenocarcinoma. *Cancer Cell*. 2014;25(6):735–747.
8. Steins A, et al. High-grade mesenchymal pancreatic ductal adenocarcinoma drives stromal deactivation through CSF-1. *EMBO Rep*. 2020;21(5):e48780.
9. Öhlund D, et al. Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer. *J Exp Med*. 2017;214(3):579–596.
10. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin*. 2020;70(1):7–30.
11. Von Hoff DD, et al. Increased survival in pancreatic cancer with nab-paclitaxel plus gemcitabine. *N Engl J Med*. 2013;369(18):1691–1703.
12. Conroy T, et al. FOLFIRINOX versus gemcitabine for metastatic pancreatic cancer. *N Engl J Med*. 2011;364(19):1817–1825.
13. Conroy T, et al. FOLFIRINOX or gemcitabine as adjuvant therapy for pancreatic cancer. *N Engl J Med*. 2018;379(25):2395–2406.
14. Neoptolemos JP, et al. Comparison of adjuvant gemcitabine and capecitabine with gemcitabine monotherapy in patients with resected pancreatic cancer (ESPAC-4): a multicentre, open-label, randomised, phase 3 trial. *Lancet*. 2017;389(10073):1011–1024.
15. Collisson E a, et al. Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. *Nat Med*. 2011;17(4):500–503.

16. Bailey P, et al. Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature*. 2016;531(7592):47–52.
17. Moffitt RA, et al. Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nat Genet*. 2015;47(10):1168–1178.
18. Maurer HC, et al. Experimental microdissection enables functional harmonisation of pancreatic cancer subtypes. *Gut*. 2019;68(6):1034–1043.
19. Funel N, Giovannetti E, Pollina LE, del Chiaro M, Mosca F, Boggi UCD. Critical role of laser microdissection for genetic, epigenetic and proteomic analyses in pancreatic cancer. *Expert Rev Mol Diagn*. 2011;Sep(7):695–701.
20. Braakman RBH, et al. Optimized nLC-MS workflow for laser capture microdissected breast cancer tissue. *J Proteomics*. 2012;75(10):2844–2854.
21. Aebersold R, Mann M. Mass-spectrometric exploration of proteome structure and function. *Nature*. 2016;537(7620):347–355.
22. Damhofer H, et al. Establishment of patient-derived xenograft models and cell lines for malignancies of the upper gastrointestinal tract. *J Transl Med*. 2015;13(1):1–14.
23. Wang D, et al. A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Mol Syst Biol*. 2019;15(2):1–16.
24. Hayashi A, et al. A unifying paradigm for transcriptional heterogeneity and squamous features in pancreatic ductal adenocarcinoma. *Nat Cancer*. 2020;1(1):59–74.
25. Peng J, et al. Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma. *Cell Res*. 2019;29(9):725–738.
26. Winter C, et al. Google goes cancer: improving outcome prediction for cancer patients by network-based ranking of marker genes. *PLoS Comput Biol*. 2012;8(5):e1002511.
27. Cancer Genome Atlas Research Network. Integrated genomic characterization of pancreatic ductal adenocarcinoma. *Cancer Cell*. 2017;32(2):185–203.e13.
28. Yang S, et al. A novel MIF signaling pathway drives the malignant character of pancreatic cancer by targeting NR3C2. *Cancer Res*. 2016;76(13):3838–3850.
29. Moore MJ, et al. Erlotinib plus gemcitabine compared with gemcitabine alone in patients with advanced pancreatic cancer: a phase III trial of the National Cancer Institute of Canada Clinical Trials Group. *J Clin Oncol*. 2007;25(15):1960–1966.
30. Markosyan N, et al. Tumor cell-intrinsic EPHA2 suppresses anti-tumor immunity by regulating PTGS2 (COX-2). *J Clin Invest*. 2019;129(9):3594–3609.
31. Amato KR, et al. Genetic and pharmacologic inhibition of EPHA2 promotes apoptosis in NSCLC. *J Clin Invest*. 2014;124(5):2037–2049.
32. Takadate T, et al. Nm23/nucleoside diphosphate kinase-A as a potent prognostic marker in invasive pancreatic ductal carcinoma identified by proteomic analysis of laser micro-dissected formalin-fixed paraffin-embedded tissue. *Clin Proteomics*. 2012;9(1):8.
33. Leca J, et al. Cancer-associated fibroblast-derived annexin A6+ extracellular vesicles support pancreatic cancer aggressiveness. *JCI*. 2016;126(9):1–17.
34. Nones K, et al. Genome-wide DNA methylation patterns in pancreatic ductal adenocarcinoma reveal epigenetic deregulation of SLIT-ROBO, ITGA2 and MET signaling. *Int J Cancer*. 2014;135(5):1110–1118.
35. Tian C, et al. Proteomic analyses of ECM during pancreatic ductal adenocarcinoma progression reveal different contributions by tumor and stromal cells. *Proc Natl Acad Sci U S A*. 2019;116(39):19609–19618.
36. Blum W, Schwaller B. Calretinin is essential for mesothelioma cell growth/survival in vitro: a potential new target for malignant mesothelioma therapy? *Int J Cancer*. 2013;133(9):2077–2088.
37. Taliano RJ, et al. Calretinin expression in high-grade invasive ductal carcinoma of the breast is associated with basal-like subtype and unfavorable prognosis. *Hum Pathol*. 2013;44(12):2743–2750.
38. Cui C, Merritt R, Fu L, Pan Z. Targeting calcium signaling in cancer therapy. *Acta Pharm Sin B*. 2017;7(1):3–17.
39. Jia D, et al. A COL11A1-correlated pan-cancer gene signature of activated fibroblasts for the prioritization of therapeutic targets. *Cancer Lett*. 2016;382(2):203–214.
40. Duxbury MS, Ito H, Zinner MJ, Ashley SW, Whang EE. EphA2: a determinant of malignant cellular behavior and a potential therapeutic target in pancreatic adenocarcinoma. *Oncogene*. 2004;23(7):1448–1456.
41. Caponi S, et al. The good, the bad and the ugly: a tale of miR-101, miR-21 and miR-155 in pancreatic intraductal papillary mucinous neoplasms. *Ann Oncol*. 2013;24(3):734–741.
42. Piersma SR, Warmoes MO, de Wit M, de Reus I, Knol JC, Jiménez CR. Whole gel processing procedure for GeLC-MS/MS based proteomics. *Proteome Sci*. 2013;11(1):17.
43. Piersma SR, et al. Feasibility of label-free phosphoproteomics and application to base-line signaling of colorectal cancer cell lines. *J Proteomics*. 2015;127(Pt B):247–258.
44. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol*. 2008;26(12):1367–1372.
45. Vizcaíno JA, et al. 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res*. 2016;44(D1):D447–D456.
46. Dijk F, et al. Unsupervised class discovery in pancreatic ductal adenocarcinoma reveals cell-intrinsic mesenchymal features and high concordance between existing classification systems. *Sci Rep*. 2019;10(1):337.
47. Viechtbauer W. Conducting meta-analyses in R with the metafor. *J Stat Softw*. 2010;36(3):1–48.
48. Le Large TYS, et al. Proteomic analysis of gemcitabine-resistant pancreatic cancer cells reveals that microtubule-associated protein 2 upregulation associates with taxane treatment. *Ther Adv Med Oncol*. 2019;11:1758835919841233.
49. García-Pravia C, et al. Overexpression of COL11A1 by cancer-associated fibroblasts: clinical relevance of a stromal marker in pancreatic cancer. *PLoS One*. 2013;8(10):1–13.
50. Thapa B, et al. Calretinin but not caveolin-1 correlates with tumour histology and survival in malignant mesothelioma. *Pathology*. 2016;48(7):660–665.
51. Furukawa T, Duguid W, Rosenberg L, Viallet J, Galloway D, Tsao M. Long-term culture and immortalization of epithelial cells from normal adult human pancreatic ducts transfected by the E6E7 gene of human papilloma virus 16. *Am J Pathol*. 1996;148(6):1763–1770.

52. Sciarrillo R, et al. Splicing modulation as novel therapeutic strategy against diffuse malignant peritoneal mesothelioma. *EBioMedicine*. 2019;39:215–225.
53. Ebbing EA, et al. Esophageal adenocarcinoma cells and xenograft tumors exposed to Erb-b2 receptor tyrosine kinase 2 and 3 inhibitors activate transforming growth factor beta signaling, which induces epithelial to mesenchymal transition. *Gastroenterology*. 2017;153(1):63–76.e14.
54. Ritchie ME, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.
55. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545–15550.
56. Luo W, Friedman MS, Shedden K, Hankenson KD, Woolf PJ. GAGE: generally applicable gene set enrichment for pathway analysis. *BMC Bioinformatics*. 2009;10:1–17.
57. Li P, et al. FHL3 promotes pancreatic cancer invasion and metastasis through preventing the ubiquitination degradation of EMT associated transcription factors. *Aging (Albany NY)*. 2020;12(1):53–69.
58. Bhasin MK, et al. Meta-analysis of transcriptome data identifies a novel 5-gene pancreatic adenocarcinoma classifier. *Oncotarget*. 2016;7(17):23263–23281.
59. Hata H, Tatemichi M, Nakadate T. Involvement of annexin A8 in the properties of pancreatic cancer. *Mol Carcinog*. 2014;53(3):181–189.
60. Iovanna J, Dusetti N. Speeding towards individualized treatment for pancreatic cancer by taking an alternative road. *Cancer Lett*. 2017;410:63–67.
61. Castillo J, et al. Surfaceome profiling enables isolation of cancer-specific exosomal cargo in liquid biopsies from pancreatic cancer patients. *Ann Oncol*. 2018;29(1):223–229.
62. Balasenthil S, et al. A migration signature and plasma biomarker panel for pancreatic adenocarcinoma. *Cancer Prev Res (Phila)*. 2011;4(1):137–149.
63. Capello M, et al. Carboxylesterase 2 as a determinant of response to irinotecan and neoadjuvant FOLFIRINOX therapy in pancreatic ductal adenocarcinoma. *J Natl Cancer Inst*. 2015;107(8):djv132.
64. Ran Y, et al. Derlin-1 is overexpressed on the tumor cell surface and enables antibody-mediated tumor targeting therapy. *Clin Cancer Res*. 2008;14(20):6538–6545.
65. Kakkar AK, Lemoine NR, Scully MF, Tebbutt S, Williamson RC. Tissue factor expression correlates with histological grade in human pancreatic cancer. *Br J Surg*. 1995;82(8):1101–1104.
66. Yao H, et al. Glypican-3 and KRT19 are markers associating with metastasis and poor prognosis of pancreatic ductal adenocarcinoma. *Cancer Biomarkers*. 2016;17(4):397–404.
67. Yokoyama S, et al. Predicted prognosis of patients with pancreatic cancer by machine learning. *Clin Cancer Res*. 2020;26(10):2411–2421.
68. Abate-Daga D, et al. A novel chimeric antigen receptor against prostate stem cell antigen mediates tumor destruction in a humanized mouse model of pancreatic cancer. *Hum Gene Ther*. 2014;25(12):1003–1012.
69. Mitsui Y, et al. Upregulation of mobility in pancreatic cancer cells by secreted S100A11 through activation of surrounding fibroblasts. *Oncol Res*. 2019;27(8):945–956.
70. Sharen G, Peng Y, Cheng H, Liu Y, Shi Y, Zhao J. Prognostic value of GLUT-1 expression in pancreatic cancer: results from 538 patients. *Oncotarget*. 2017;8(12):19760–19767.
71. Sun H, Dai X, Han B. TRIM29 as a novel biomarker in pancreatic adenocarcinoma. *Dis Markers*. 2014;2014:317817.

Supplementary Appendix Chapter 4

Microdissected pancreatic cancer proteomes reveal tumor heterogeneity and therapeutic targets

Supplemental materials and methods

Sample collection and processing

Approval from the Local Medical Ethical committee at the Amsterdam UMC, location VU University Medical Center was received for the Biobank (#14038). All patients gave informed consent for tissue sampling, clinical data analysis and molecular analysis. Consecutive snap-frozen tumor samples from January 2014 until November 2015 from the VU University Medical Center were evaluated for their quality and tumor percentage. The workflow is described in Figure 1A. After pathological evaluation, 16 samples were eligible for further LCM analysis and prepared as follows: frozen slides of 10µm thickness were applied on PEN foil slides (Leica, Germany). Sections were stained for 1 min with Mayers Hematoxylin, rinsed in RNAse free sterile water, and dehydrated in sequential concentrations of Ethanol and 100% Xylene. Dehydrated slides were stored at -80°C and thawed once for the LCM procedure. Additionally, unseparated bulk tumor of 11 matched and 5 unmatched samples were prepared for protein identification. This yielded a total of 21 PDAC samples.
Clinical parameters were collected prospectively and overall survival (OS; defined by registered death of the patient, or last visit to the clinic) and disease-free survival (DFS; defined as disease-free period between resection and registered recurrence of disease) were annotated. Two patients were censored for OS analysis and the differential analysis with short versus long survival, since they succumbed of complications after surgery, defined as a mortality within 60 days after surgery. Five patients were censored for DFS analysis due to different reasons; lost to follow up (n=1), mortality due to surgery (n=2), R2 resection (n=1), metastatic disease at time of resection (n=1). One patient in the cohort had stage IV disease. This patient was preoperatively predicted to suffer from duodenal adenocarcinoma, but pathological review showed PDAC. Since the liver metastasis was resected synchronously, the patient was not taken along for differential analysis comparing OS/DFS.

Laser capture microdissection procedure

LCM was performed on the Leica LMD7000 instrument (Leica-Microsystems, Wetzlar, Germany), for a total surface of 3x106 µm2 per compartment. Selected areas were captured in 0.1% RapiGest SF Surfactant (Waters, Milford, MA) and stored until further preparation. Samples were sonicated and reduced to a final concentration of 5mM dithiothreitol (DTT, Sigma-Aldrich, Saint Louis, MO) and 15mM iodoacetamide (IAA, Sigma-Aldrich, Saint Louis, MO). Sequencing-grade modified trypsin (Promega, Madison, WI) was added overnight to a final concentration of 7ng/µl. Digestion was stopped with acidification by trifluoroacetic acid (TFA). The peptide mixture was centrifuged and supernatant was transferred to a glass-lined MS/MS auto sampler vials. Samples were brought up to 20µl volume and stored at -80°C until further analysis.

Peptide preparations of PDX models and bulk tumor

Animal work was performed in a previous study according to protocols approved by the animal experiment ethical committee at the Amsterdam UMC (location AMC, protocol DTB102348, LEX102774). NSG (Cg-Prkdcscid Il2rgtm1Wjl/SzJ) mice used for the experiments were bred in-house. Frozen tumors from 10 PDX models[1] were used for proteome analysis. A minimum of 100 mg of tissue was digested in lysis buffer (9M urea, 20mM HEPES pH 8.0, 1mM Na3VO4, 2.5mM Na4P2O7, 1mM Na2C3H7PO6). 50µg of protein was loaded on a NuPAGE 4-12% gradient gel (Invitrogen, Carlsbad, CA). Proteins were digested and extracted according our whole-in-gel protocol described previously[2]. In short, proteins were separated by gel-electrophoresis, fixed and stained with Coomassie Brilliant Blue solution. After reduction and alkylation, proteins were digested overnight with trypsin and extracted from the gel for further analysis. Bulk tumors were lysed and in-solution digestion was performed according to our laboratories protocol after reduction and alkylation(3). Digestion was inhibited by acidification with TFA and peptide eluates were desalted with 20 µl StageTips and peptides were stored in glass-lined autosampler vials[3] until measurement.

Nano-LC-MS/MS

Peptides were separated by an Ultimate 3000 nanoLC-MS/MS system (Dionex LC- Packings, Amsterdam, The Netherlands) equipped with a 40 cm × 75 µm ID fused silica column custom packed with 1.9 µm 120Å ReproSil Pur C18 aqua (Dr Maisch GMBH, Ammerbuch-Entringen, Germany). After injection, peptides were trapped at 6 µl/min on a 10 mm × 100 µm ID trap column packed with 5 µm 120 Å ReproSil Pur C18 aqua in 0.05% formic acid. Peptides were separated at 300 nl/min in a 10–40% gradient (buffer A: 0.5% acetic acid (Fisher Scientific, the Netherlands), buffer B: 80% ACN, 0.5% acetic acid) in 130 min (150 min inject-to-inject). Eluting peptides were ionized at a potential of +2 kVa into a Q Exactive HF mass spectrometer (Thermo Fisher, Bremen, Germany). Intact masses were measured at resolution 70.000 (at m/z 200) in the orbitrap using an AGC target value of 3E6 charges. The top 15 peptide signals (charge-states 2+ and higher) were submitted to MS/MS in the HCD (higher- energy collision) cell (1.6 amu isolation width, 25% normalized collision energy). MS/MS spectra were acquired at resolution 17.500 (at m/z 200) in the orbitrap using an AGC target value of 1E6 charges, a maxIT of 32 ms and an underfill ratio of 0.1%. Dynamic exclusion was applied with a repeat count of 1 and an exclusion time of 30 s.

Protein annotation and data analysis

MS/MS spectra were searched against the Swissprot FASTA file (LCM data: release march 2017, 42161 entries, canonical and isoforms.

PdX data:uniprot_human_referenceproteome_2014_01_NO_fragments_42104entries.fasta (61552 entries);

Uniprot_Mus_musculus_reference_proteome_2015_06_NO_FRAGMENTS_Canonic al and isoforms _34331entries.fasta (42296 entries)) using MaxQuant[4] (version 1.5.8.0). Enzyme

specificity was set to trypsin and up to two missed cleavages were allowed. Cysteine carboxamidomethylation (Cys, +57.021464 Da) was treated as fixed modification and methionine oxidation (Met,+15.994915 Da) and N-terminal acetylation (N-terminal, +42.010565 Da) as variable modifications. Peptide precursor ions were searched with a maximum mass deviation of 4.5 ppm and fragment ions with a maximum mass deviation of 20 ppm. Peptide and protein identifications were filtered at an FDR of 1% using the decoy database strategy. The minimal peptide length was 7 amino acids. Proteins that could not be differentiated based on MS/MS spectra alone were grouped to protein groups (default MaxQuant settings). Searches were performed with the label-free quantification option selected.

A match-between-runs setting was implemented for analysis of low abundant proteins in the LCM database. Protein compartment specificity was correlated to mouse and human-specific proteins identified from PDX PDAC tumors (protein had to be uniquely mouse- or human-specific). Proteins had to be identified in 5 out of 10 samples for further selection in the species-specific list to reduce heterogeneity. Mouse genes were converted to human nomenclature for GSEA (See Supplemental Data 5). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository[5] with the dataset identifier PXD011289 and PXD017393. One of the tumor samples (sample CC) showed low identifications and inadequate MS/MS data and was omitted from further analysis.

In silico validation of prognostic markers

Data of publicly available transcriptomic (microarray or RNAseq) datasets with survival data were downloaded from GEObase and each dataset was scaled to a mean of zero, with a standard deviation of 1 to allow meta-analysis. Univariate cox proportional hazard regression models were evaluated for genes of interest. The Metafor R package[6] was used to perform meta-analysis validation of identified prognostic markers. Combined hazard risk of proteins of interest and confidence intervals are reported if genes were identified in a minimum of 3 out of the 7 datasets.

Immunohistochemistry (IHC) validation of tissue microarrays (TMA)

Proteins of interest in both compartments (COL11A1, CALB2) were evaluated in two independent cohorts (N=95 and N=95) by staining TMAs of resected patients treated at the Pisa University Hospital. Written informed consent was obtained from all patients, and ethical approval was provided by the Ethics Committee of the Pisa University hospital (Pisa, Italy, date of approval: July 3, 2013 (file number 3909)). The PDAC TMA has been detailed previously[7] and represents an independent, non-overlapping cohort of patients. Assessment of IHC staining of COL11A1 was performed in 3 tumor cores containing representatives regions of the desmoplastic reaction, and the expression was evaluated in relation to the stromal surface, as described previously[8]. The cores with <1% positive staining were assigned a score of 0, while cores with positive staining between 1 and 10% were assigned a score of 1, cores with positive staining between 10% and 50% were assigned a score of 2, and cases with more than 50%

positive staining were assigned a score of 3. Total variation in staining was defined by multiplying the number of positive cores (0-3) by the field's staining intensity (0-3), yielding a total score of 0-9. Samples were defined as "high COL11A1", when staining score was > 5; and "low COL11A1" when staining score was ≤ 5. Scoring for CALB2 was performed taking into account both the percentage of neoplastic cells stained and also the intensity of the staining (marked as 1+/2+/3+), as reported previously[9]. The IHC score for each core was calculated separately and then an average of 3 cores was taken as the final score for each sample/patient. Samples were then categorised according to the median value and were defined as "high CALB2" when staining score was>median; and "low CALB2" when staining score ≤median.
The immunostaining was double-blind scored and correlated to clinical data by two researchers (NF and EG).

Western blot validation of EPHA2 expression

PANC1 (ATCC,Manassas, WI) was cultured in RPMI medium (Lonza, Switzerland). Capan-2 (ATCC, Manassas, WI) and Hs766t (ATCC, Manassas, WI) were cultured in DMEM medium (Lonza). Media was supplemented with 10% heat-inactivated fetal calf serum (Biowest, France) and 1% penicillin and streptomycin (Lonza). An immortalized pancreatic ductal cell line (HPDE, kindly supplied by dr. Tsao, Ontario, Canada[10]) was cultured in supplemented KGM medium (Lonza). Cell lysates were created with diluted 10x RIPA buffer (Abcam, UK) containing protease (cOmpleteTM mini EDTA- free protease inhibitor cocktail tablets, Roche, Switzerland) and phosphatase inhibitors (1mM Na3VO4, 2.5mM Na4P2O7, 1mM Na2C3H7PO6) according to manufacturer's protocol. The BCA protein estimation method (Pierce BCA protein assay kit, ThermoFisher, Waltham, MA) was used to evaluate protein yield. 20 µg of protein was denatured with NuPAGE 4x LDS buffer (ThermoFisher, Waltham, MA), 10% DTT and heating. Separated Proteins were transferred to nitrocellulose membranes (EMDMillipore,Burlington, MA). Blocking was performed with 5% blotting- grade blocker non-fat dry milk (Biorad,Hercules, CA) in PBS and 0.1% Tween-20 (PBST). Primary antibodies were incubated in 5% BSA (Sigma-Aldrich, Saint Louis, MO) in PBST followed by secondary antibodies in 5% blocking buffer. Proteins were detected with SuperSignal West Pico Chemoluminescent substrate (ThermoFisher, Waltham, MA) and visualized by an Uvitec Imaging station (Cleaver Scientific, UK).

List of antibodies

Target	Conjugate	Antibody	Cat #	Supplier	Application
EPHA2	None	EphA2 (D4A2) XP Rabbit mAb	#6997	Cell Signaling Technology	WB
Phospho-EPHA2	None	Phopsho-EphA2 (Tyr588) (D7X2L)	#12677	Cell Signaling Technology	WB
COL11A1	None	DMTX invascan	DMTX invascan	Oncomatryx	IHC
CALB	None	Anti-Calretinin mAb	Ab702	Abcam	IHC
Anti-rabbit	HRP	Anti-rabbit IgG, HRP-linked Antibody	#7074	Cell Signaling Technology	WB
Anti-rabbit	HRP	Goat-Anti-Rabbit IgG H&L	Ab205718	Abcam	IHC

Stable EPHA2 knockdown

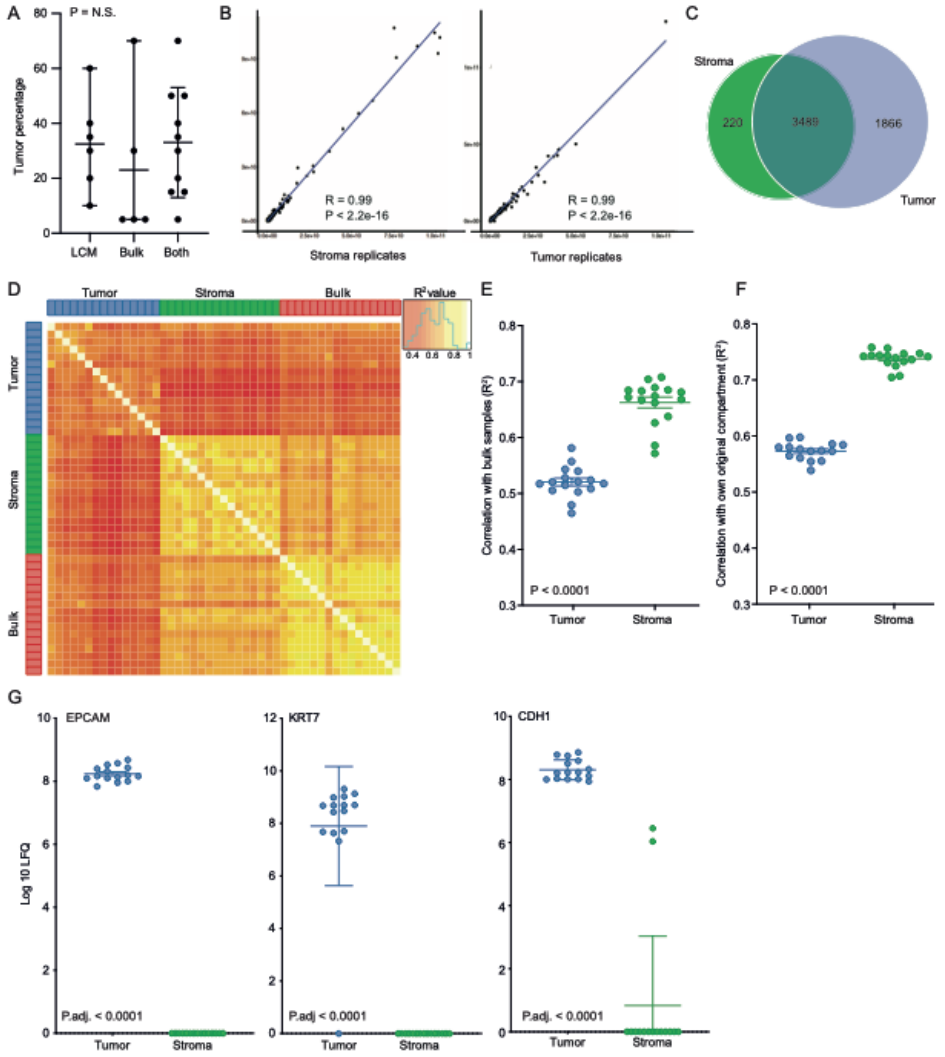
Lentiviral plasmids were produced by transfecting HEK293T cells with *EPHA2* targeting pLKO.1 constructs (MISSION shRNA Library clone numbers TRCN0000006403 and TRCN0000197131) and a non-targeting sequence negative control (shc002). Transfected supernatant was collected after 48 hours and filtered through a 0.45µm filter (EMDMillipore, Burlington, MA). At 30% confluency, PDAC cell lines Capan-2 and Hs766t were transduced and subsequently selected after 48 hours with 2 µg/ml puromycin (Sigma-Aldrich, Saint Louis, MI). Knockdown efficiency was evaluated by Western blot as described above.

In vitro validation of drug target

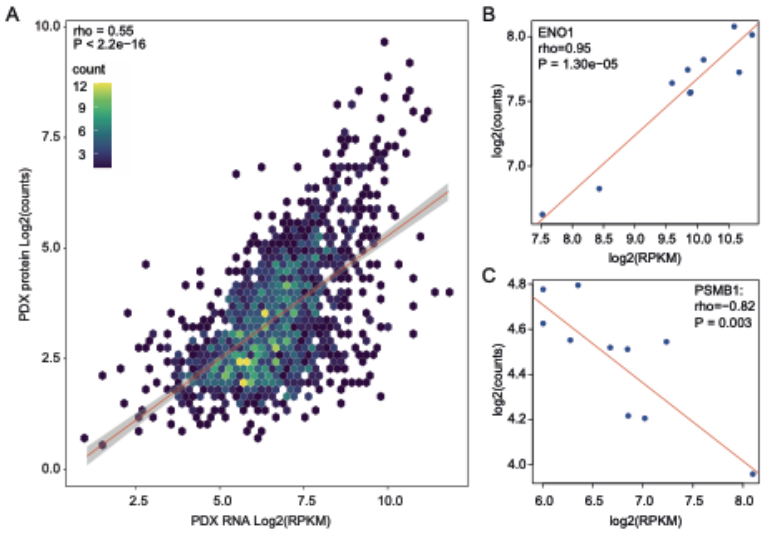
Cells were plated in 96-wells plates and allowed to attach overnight (PANC1 3000/well, Capan-2 5000/well, Hs766t 5000/well, HPDE 5000/well). Growth was evaluated over 72 hours in respect to the control at the start of the experiment. To evaluate proliferation, the doubling time was calculated. A specific EPHA2 inhibitor ALW-II-41-27 (APExBIO, Houston, TX) was evaluated for cytotoxic effect. DMSO was used as control. Effect on cell proliferation was quantified with Sulforhodamine B (SRB, Sigma Aldrich, Saint Louis, MO) staining of protein and was subsequently measured for absorbance at 492nm in the Synergy microplate reader (Biotek Cytation3 Cell Imaging Multi-Mode Reader, Waltham, MA) as described previously[11]. Migration was evaluated by transwell migration after staining cells with fluorescent dye as described before[12]. As attractant, FCS 1% or 1µg/mL EGF was used. Cells were pretreated with ALW-II-41-27 for 15 min prior to migration. Migration was evaluated every 2 minutes for 3 hours and values were controlled for background, and values from no-attractant controls were subtracted at each time point.

Statistical analysis

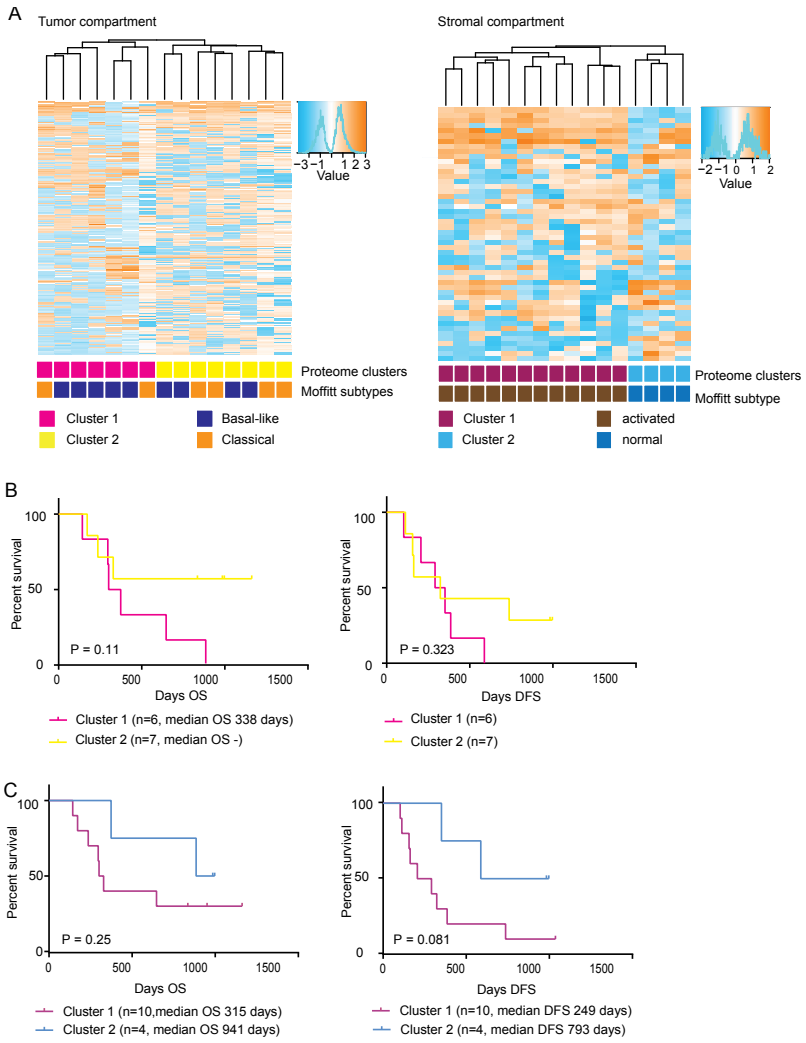
Data was analyzed with R (version 3.5.2.). Zeros were imputed based on normal distribution standard deviation of log transformed intensity data[13] before differential expression analysis. The highest measured intensity was used for further analysis from replicates. Unsupervised clustering was performed on z-score normalized data and Euclidean distance. Differential compartment expression was tested with paired statistics analysis (Limma R package)[14] and corrected for multiple testing by the Benjamini Hochberg equation. Differential expression between tumor and bulk tissue was performed by unpaired Limma statistics. Gene set analysis[15] was performed in R. Prognostic proteins were identified by unpaired limma statistics of a group comparison (short OS, <1 year versus longer OS, >2 years). Technical and biological replicates were evaluated with Pearson's correlation. Group comparison of correlations was tested with unpaired student's t-test. *In vitro* experimental comparisons were evaluated by paired or unpaired two-tailed student's t-test. For the migration assay, the AUC was calculated and tested with the Welch's t-test. Complete clinicopathological, follow-up, and recurrence data were available from prospectively maintained databases. Correlation of clinicopathological characteristics and gene panel expression (based on z-score group selection) with DFS/OS were evaluated with Kaplan-Meier curves and the Log- rank test. Prognostic value of IHC scoring was tested with uni- and multivariate analysis. Error-bars show the mean ± SEM. A *P* value of <.05 was considered statistically significant.



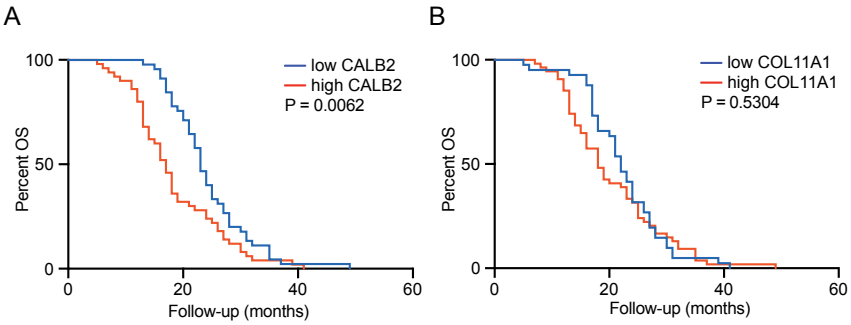
Supplemental Figure 1. Quality control of samples and LC-MS/MS data. A. Dot plot of tumor purity of samples used for analysis, LCM only (n=6), bulk only (n=5), both (n=10) (P = ns, two-way unpaired t-test between all groups). B. Quality control with technical replicates shows high Pearson Coefficient correlation of replicates of tumor and stroma samples (R2 = 0.99, P < 0.0001). C. Venn diagram of tumor and stromal proteins identified in each compartment. D. Heatmap of inter-sample correlation analysis. E. Bulk samples (n=16) correlate significantly better with stromal samples than tumor samples (two-way unpaired t-test P < 0.0001). F. Inter- sample correlation shows high inter-patient correlation in stroma compared to tumor samples (two-way unpaired t-test, P<0.0001). Data represent mean and SD. G. Dot plot showing specific tumor marker expression of EPCAM /KRT7/CDH1 in tumor and stromal areas (both n = 15) with low to none expression in stroma, indicating minimal contamination. Data represents mean and SD (paired-limma test, adjusted for multiple testing, P<0.0001).



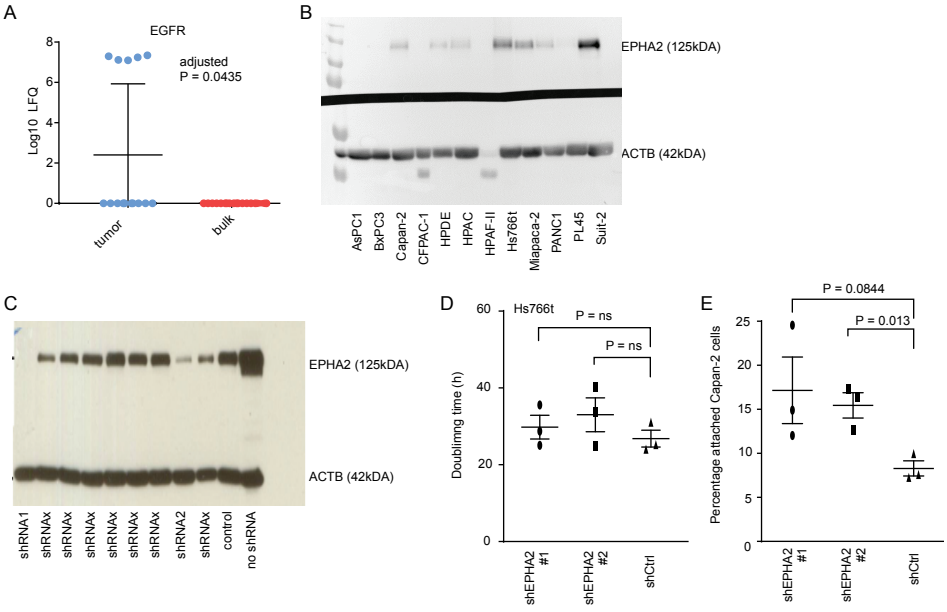
Supplemental Figure 2. Correlation between proteomics and transcriptomics in PDAC PDXs. A. Scatter plot shows correlation for transcriptomics and proteomics data of 10 PDX models for PDAC. Data are visualized in bins and color count shows the frequency of genes for each bin. Regression line in red is estimated using the built-in lm() function in R. **B.** Correlation analysis of Enolase 1 (ENO1) protein counts with transcript levels. Dots denote individual PDXs. **C.** As panel B, for Proteasome 20S Subunit Beta 1(PSMB1) protein.



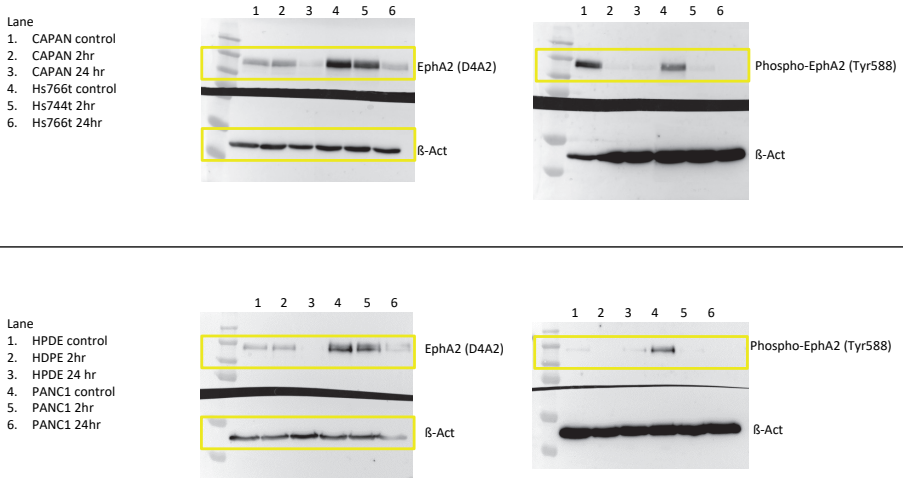
Supplemental Figure 3. Clustering to evaluate proteome subtypes. A. Protein clustering on most variable genes (top 500 proteins in tumor samples, top 50 proteins in stromal samples) showed two protein subclasses on either compartment. Samples were associated to known classifiers(15,17) by Z-score ranking. Correlation to known subtypes shows good correlation to previous established stromal subtypes. **B.** Kaplan-Meier curves of OS and DFS from proteome cohort of epithelial samples (log-rank test, P = 0.11 and P = 0.323 respectively) and stroma (**C**) (log-rank test, P = 0.25 and P = 0.08 respectively for OS and DFS).



Supplemental Figure 4. Validation of prognostic markers in a second independent TMA cohort. A. Kaplan-Meier curves of high (red, n=45) or low (blue, n=50) expression of CALB2 in a second cohort shows significant correlation to OS (P < 0.009, median survival 17 versus 23 months, log-rank test). **B.** Kaplan-Meier curves of high (red, n=41) or low (blue, n=54) expression of COL11A1 shows no significant correlation the second validation cohort (P = ns, median survival 18 versus 22 months, log-rank test).



Supplemental Figure 5. EPHA2 inhibition as target in PDAC A. Expression of EGFR in bulk (n=16) and tumor (n=15) samples (limma-test corrected for multiple testing, adjusted P < 0.05). Data represents mean with SD. **B.** Western blot analysis of PDAC cell lines shows variance in EPHA2 expression. **C.** shRNA knock-down of EPHA2 control by Western Blot. shRNA 1 and shRNA 2 were chosen or further experiments. shRNAx were identified as non-functional shRNAs. **D.** Doubling time evaluation of Hs766t (P = non-significant, unpaired two-way t-test). Data represents mean with SEM (n = 3).**E.** Evaluation of detachment of Capan-2 with shRNAs against EPHA2 upon trypsinization. Cells were plated and attached overnight and detached by 1% trypsin. At time of evaluation, number of detached cells were counted and normalized to total number of cells. shRNA clone 2 detached slower (P = 0.013, unpaired two-way t-test). Data represents mean with SEM (n = 3).



Supplemental Figure 6. Full unedited western blots of Figure 5D.

Supplemental Table 1. Additional information for cohort.

	N	Median (SE)	*P*-value
Overall Survival	19	369±189	
Age			
<69	9	353±170	*0.195*
≥69	10	369±248	
Sex			
Female	10	353±51	*0.241*
Male	9	581±385	
Disease stage			
Stage I-II	11	637±241	*0.846*
Stage III-IV	8	323±46	
Lymph nodes			*0.946*
N0	5	353±67	
N1	8	581±222	
N2	6	296±80	
Adjuvant therapy			
No adjuvant therapy	8	292±40	*0.040**
Gemcitabine	11	730±277	

Supplemental Table 2. Clinicopathological information with samples.

Patient	Age at diagnosis	Sex	Tumor differentiation	T-Stage	N-Stage	M-stage	R	Stage (AJCC 8th edition)	Tumor %	Adjuvant therapy	Recurrence	DFS (days)	Survival	OS (days)
PDAC1	58	F	Intermediate	2	1	0	1	2B	5-10%	YES	PD	529	D	740
PDAC2	65	F	Intermediate	3	0	0	1	2A	50%	YES	PD	595	D	885
PDAC3	78	M	Intermediate	2	2	0	0	3	40%	NO	PD	296	D	301
PDAC4	54	M	poor	2	2	0	2	3	5-10%	NO	UN		D	243
PDAC5	50	F	poor	3	0	0	1	2A	5-10%	PARTLY	PD	105	D	143
PDAC6	77	F	poor	3	0	0	1	2A	15%	NO	PD	210	D	296
PDAC7	72	F	Intermediate	3	1	0	1	2B	70%	PARTLY	PD	385	D	647
PDAC8	69	F	poor	2	2	0	1	3	60%	YES	PD	351	D	374
PDAC9	70	M	well	3	1	0	1	2B	30%	YES	PD	737	A	1227
PDAC10	80	M	Intermediate	2	1	0	1	2B	20%	NO	UN		D, excluded	20
PDAC11	63	F	Intermediate	2	0	0	1	1B	35%	PARTLY	PD	339	D	358
PDAC12	73	M	Intermediate	2	2	0	1	3	35%	YES	NR		A	1217
PDAC13	60	M	poor	2	1	0	0	2B	20%	YES	NR		A	1118
PDAC14	75	M	Intermediate	1	0	0	1	1A	50%	NO	NR		A	1037
PDAC15	78	F	poor	2	1	0	1	2B	5-10%	NO	PD	235	D	309
PDAC16	67	M	well	3	2	0	0	3	10%	NO	PD	159	D	236
PDAC17	64	M	poor	2	1	0	1	2B	35%	NO	UN		D, excluded	31
PDAC18	73	M	poor	2	1	0	1	2B	40%	PARTLY	PD	167	D	329
PDAC19	71	M	Intermediate	3	2	0	1	3	35%	YES	PD	329	A	923
PDAC20	67	F	Intermediate	2	1	0	1	2B	15%	NO	PD	122	D	172
PDAC21	67	M	poor	2	1	1	0	4	70%	NO	PD	335	D	589

PD: Progressive disease NR: no recurrence A: alive at last follow up Adjuvant therapy: yes= 6 cycles total regimen, <6 cycles=partly, no= 0 cycles

UN: unknown D: deceased Excluded: death due to complication, excluded from survival analysis

Supplemental Table 3. Clinical characteristics of two independent PDAC cohorts for TMAs. Multivariate analysis (when P <= 0.1 at univariate).

Univariate analysis	N (%)	OS months mean (95% CI)	*p*-value	N(%)	OS months mean (95% CI)	*p*-value
No. patients	95 (100)	20.4 (18.0-22.8)		95 (100)	21.4 (19.6-22.9)	
Age, year – at time of diagnosis			*0.339*			*0.359*
>65	36 (37.9)	19.06 (16.1–22.0)		41 (43.2)	22.7 (19.7-24.6)	
≤65	59 (62.1)	21.22 (17.8-24.6)		54 (56.8)	20.5 (18.4-22.7)	
Gender			*0.016*			*0.476*
Female	42 (44.2)	23.69 (19.7-27.7)		44 (46.3)	20.1 (17.6-22.7)	
Male	53 (55.8)	17.79 (15.1-20.5)		51 (53.7)	22.5 (20.7-24.4)	
Vascular infiltration			*0.012*			*0.065*
No	33 (34.7)	24.7 (19.9-29.5)		45 (47.4)	22.8 (20.2-25.4)	
yes	62 (65.3)	18.1 (15.6-20.6)		50 (52.6)	19.9 (17.9-21.8)	
Tumor grade			*0.033*			*0.054*
Grade 1-2	48 (51.5)	22.8 (19.1-26.7)		37 (38.9)	23.1 (20.2-26.4)	
Grade 3	47 (49.5)	17.9 (15.2-20.6)		58 (61.1)	20.0 (18.2-21.9)	
Resection margin			*0.208*			*0.245*
No	50 (52.6)	15.06 (12.0-18.1)		56 (58.9)	22.4 (20.5-24.4)	
Yes	45 (47.4)	21.66 (18.6-24.7)		58 (41.1)	19.2 (16.8-22.2)	
COL11A1			*0.016*			*0.530*
low	50 (52.6)	23.0 (19.3-26.7)		41 (43.2)	22.3 (10.1-24.5)	
high	45 (47.4)	17.5 (18.0-22.8)		53 (56.8)	20.4 (18.1-22.7)	
CALB2			*0.009*			*0.006*
Low	47 (49.5)	23.6 (19.8-27.3)		45 (47.4)	24.3 (22.3-26.4)	
high	48 (51.5)	17.3 (14.5-20.0)		50 (52.6)	18.5 (16.2-20.7)	
Multivariate analysis	**df**	**Risk of death, HR (95% CI)**	***p*-value**	**df**	**Risk of death, HR (95% CI)**	***p*-value**
Gender (female vs. male)		0.53	*0.105*	/	/	/
Grading (3 vs. 1-2)		0.57	*0.037*		0.68 (0.33-1.55)	*0.108*
Vascular infiltration (no vs. yes)		0.68	*0.144*		0.59 (0.27-1.19)	*0.084*
COL11A1 (low vs. high)		0.92	*0.312*	/	/	/
CALB2 (low vs. high)		0.32	*0.004*		0.36 (0.25-0.67)	*0.008*

Additional Supplemental Data

Supplemental Data 1: Sample data LCM resource

Supplemental Data 2: Protein data LCM resource

Supplemental Data 3: Sample data PDX resource

Supplemental Data 4: Protein data PDX resource

Supplemental Data 5: List of mouse/human specific proteins identified from PDX

Supplemental Data 6: Differential analysis tumor/stromal compartment

Supplemental Data 7: Differential analysis on short/long survival stroma

Supplemental Data 8: Differential analysis on short/long survival tumor

References supplemental methods

1. Damhofer, H. *et al.* Establishment of patient-derived xenograft models and cell lines for malignancies of the upper gastrointestinal tract. *J. Transl. Med.* **13**, 115 (2015).
2. Piersma, S. R. *et al.* Whole gel processing procedure for GeLC-MS/MS based proteomics. *Proteome Sci.* **11**, 17 (2013).
3. Piersma, S. R. *et al.* Feasibility of label-free phosphoproteomics and application to baseline signaling of colorectal cancer cell lines. *Proteome Quest Understand Biol. Dis. HUPO 2014* **127**, 247–258 (2015).
4. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367 (2008).
5. Vizcaíno, J. A. *et al.* ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* **32**, 223–226 (2014).
6. Viechtbauer, W. Conducting Meta-Analyses in *R* with the **metafor** Package. *J. Stat. Softw.* **36**, (2010).
7. Le Large, T. Y. S. *et al.* Proteomic analysis of gemcitabine-resistant pancreatic cancer cells reveals that microtubule-associated protein 2 upregulation associates with taxane treatment. *Ther. Adv. Med. Oncol.* **11**, 175883591984123 (2019).
8. García-Pravia, C. *et al.* Overexpression of COL11A1 by Cancer-Associated Fibroblasts: Clinical Relevance of a Stromal Marker in Pancreatic Cancer. *PLoS ONE* **8**, e78327 (2013).
9. Thapa, B. *et al.* Calretinin but not caveolin-1 correlates with tumour histology and survival in malignant mesothelioma. *Pathology (Phila.)* **48**, 660–665 (2016).
10. Furukawa, T. *et al.* Long-term culture and immortalization of epithelial cells from normal adult human pancreatic ducts transfected by the E6E7 gene of human papilloma virus 16. *Am. J. Pathol.* **148**, 1763–1770 (1996).
11. Sciarrillo, R. *et al.* Splicing modulation as novel therapeutic strategy against diffuse malignant peritoneal mesothelioma. *EBioMedicine* **39**, 215–225 (2019).
12. Ebbing, E. A. *et al.* Esophageal Adenocarcinoma Cells and Xenograft Tumors Exposed to Erb-b2 Receptor Tyrosine Kinase 2 and 3 Inhibitors Activate Transforming Growth Factor Beta Signaling, Which Induces Epithelial to Mesenchymal Transition. *Gastroenterology* **153**, 63-76.e14 (2017).
13. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47–e47 (2015).
14. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* **102**, 15545 (2005).

PART TWO

INNOVATIVE ANALYSES OF miRNAs AND mRNAs

Chapter 5

New avenues in pancreatic cancer: exploiting microRNAs as predictive biomarkers and new approaches to target aberrant metabolism

Capula M, Mantini G, Funel N, Giovannetti E

Expert Review for Clinical Pharmacology, 2019

EXPERT REVIEW OF CLINICAL PHARMACOLOGY
2019, VOL. 12, NO. 12, 1081–1090
https://doi.org/10.1080/17512433.2019.1693256

Taylor & Francis
Taylor & Francis Group

SPECIAL REPORT

New avenues in pancreatic cancer: exploiting microRNAs as predictive biomarkers and new approaches to target aberrant metabolism

Mjriam Capula[a], Giulia Mantini[a,b], Niccola Funel[a] and Elisa Giovannetti[a,b]

[a]Cancer Pharmacology Lab, AIRC Start-Up Unit, Fondazione Pisa per la Scienza Pisa, Pisa, Italy; [b]Department of Medical Oncology, VU University Medical Center, Cancer Center Amsterdam, Amsterdam, Netherlands

ABSTRACT

Introduction: Most pancreatic cancer patients are diagnosed at advanced-stages and first-line regimens (FOLFIRINOX and gemcitabine/nab-paclitaxel) provide limited survival advantage and are associated with considerable toxicities. In this grim scenario, novel treatments and biomarkers are warranted.

Areas covered: MicroRNAs (miRNAs) emerged as biomarkers for cancer prognosis and chemoresistance and blood-based miRNAs are being evaluated as indicators of therapeutic activity. Moreover, aberrant metabolism, such as aerobic glycolysis, has been correlated to tumor aggressiveness and poor prognosis. Against this background, innovative approaches to tackle metabolic aberrations are being implemented and glycolytic inhibitors targeting lactate dehydrogenase-A (LDH-A) showed promising effects in preclinical models. A PubMed search was used to compile relevant publications until February 2019.

Expert opinion: Analysis of tissue/circulating miRNA might improve selection for optimal treatment regimens. For instance, miR-181a modulation seems to predict response to FOLFIRINOX. However, we need further studies to validate predictive miRNA profiles, as well as to exploit miRNAs for treatment-tailoring. Several miRNAs have also a key role in regulating metabolic aberrations. Since preliminary evidence supports the development of new agents targeting these aberrations, such as LDH-A inhibitors, the identification of biomarkers for these treatments, including the above-mentioned miRNAs, should shorten the gap between preclinical studies and personalized therapies.

ARTICLE HISTORY
Received 3 August 2019
Accepted 12 November 2019

KEYWORDS
FOLFIRINOX; gemcitabine; metabolic reprogramming; microRNAs; nab-paclitaxel pancreatic cancer; Warburg effect

1. Introduction

Pancreatic ductal adenocarcinoma (PDAC) is currently the seventh leading cause of cancer-related deaths worldwide and despite the continuous improvements in its detection and management, the 5-year survival rate still stands below 9% [1]. This dismal prognosis is mostly due to late diagnosis as well as poor efficacy of current treatments [2]. In fact, because of the lack of biomarkers for early detection and of the retroperitoneal location, which often determines the absence of signs or symptoms in the early stages, this tumor is generally diagnosed at advanced stages. Though for resectable or borderline resectable patients surgical resection is the primary treatment for PDAC, the diagnosis at advanced stages and the invasive nature of this tumor impede the potential curative resection, rendering chemotherapy the sole treatment option for patients with metastatic or locally advanced PDAC [3].

In the last few years, several novel chemotherapeutic regimes have been developed and some progress in term of clinical response has been made. However, the impact on patient overall survival (OS) is rather limited. Such failure is caused, at least in part, by inherent or acquired chemoresistance [4]. In addition, several new regimens cause a significant increase in hematologic and extra-hematologic toxicities compared to gemcitabine-alone. Thus, we urgently need studies to identify molecular biomarkers, such as microRNAs, that could predict response to therapy in order to maximize the efficacy of treatments and avoid useless side effects for non-responding patients.

More effective therapeutic approaches are also warranted and the renewed interest in tumor metabolism has generated hope that a new class of anti-cancer treatment strategies could target aggressive tumors such as PDAC.

In fact, most PDAC cells exhibit profound metabolic alterations [5]. Among these abnormalities, one of the most common is the Warburg effect, which consists in an increased glycolysis even in the presence of oxygen [6]. Therefore, compounds that inhibit components of the glycolytic pathway could represent an innovative and effective anticancer strategy.

In the present review, we summarize the main therapeutic options for PDAC, and then critically discuss the use of microRNAs as novel potential biomarkers to predict drug activity. Moreover, we reported new experimental compounds that target glycolytic metabolism, analyzing their potential use to improve current therapies against PDAC. To cover these issues a PubMed search was used to compile relevant publications, until February 2019.

CONTACT Elisa Giovannetti elisa.giovannetti@gmail.com Department of Medical Oncology, VU University Medical Center, Cancer Center Amsterdam, Amsterdam, Netherlands

Article Highlights

- Pancreatic cancer has a dismal prognosis mostly due to late diagnosis and poor efficacy of current treatments: new treatments and biomarkers are warranted
- MicroRNAs (miRNAs) are emerging as predictive biomarkers of response and should be investigated to optimize new therapeutic strategies
- Innovative approaches to tackle cancer metabolic aberrations, such as glycolytic inhibitors targeting lactate dehydrogenase-A (LDH-A), showed promising effects in preclinical models
- The parallel development of new drugs targeting LDH-A and of biomarkers for these treatments, including miRNAs, should shorten the gap between preclinical studies and personalized therapies.

2. Standard treatments

2.1. Gemcitabine and gemcitabine-based combinations

Gemcitabine is a pyrimidine analogue (2′,2′-difluorodeoxycytidine, dFdC) which exerts its antiproliferative action after conversion into active triphosphorylated nucleotides, interfering with DNA synthesis and targeting ribonucleotide reductase. It is extensively prescribed to treat a variety of other solid tumors such as pancreatic, breast, ovarian, bladder or non-small-cell lung (NSCLC) cancers [7]. Until a few years ago, gemcitabine monotherapy has been used as the first-line treatment for metastatic PDAC, since it provided an increased clinical benefit compared to 5-fluorouracil (5-FU). However, the median OS observed in metastatic PDAC patients administrated with gemcitabine monotherapy was only 5.65 months, with a very low response rate (i.e. 5.4%) [8]. For this reason, several studies were performed to improve patients' prognosis, testing the combination of gemcitabine with other treatment modalities.

Unfortunately, the combination of gemcitabine with capecitabine or cisplatin showed only a marginal improvement in term of OS, and no statistically significant differences were observed when comparing different combinations of chemotherapeutic agents to gemcitabine monotherapy [9,10].

Multiple clinical trials were also conducted to test the combination of gemcitabine with biologically targeted agents, including epidermal growth factor receptor (EGFR) tyrosine kinase inhibitors, such as erlotinib. The combination of gemcitabine with erlotinib resulted in a slight, though significant, improvement in terms of survival benefit, but it was not considered clinically relevant [11]. This might be explained by the fact that EGFR mutations, that are used to guide the treatment of NSCLC with EGFR-tyrosine kinase inhibitors [12], are extremely rare in PDAC samples [13]. Skin rash was initially proposed as a surrogate marker of efficacy, but it failed to identify patients with clinical benefit, as reported in a randomized phase II dose escalation trial [14].

The first study of a gemcitabine-combined regimen which showed clinically relevant results was the phase III trial IMPACT, evaluating the combination of nab-paclitaxel (Abraxane®) with gemcitabine as the first-line option for patients with advanced or metastatic PDAC [15]. Nab-paclitaxel is a nanoparticle albumin-bound paclitaxel and represents the first nanotechnology-based drug in cancer treatment [16]. Given the reduced diameter of these particles, nab-paclitaxel has a greater distribution volume and a faster clearance than conventional paclitaxel [17]. Moreover, since albumin-paclitaxel complexes have sizes virtually identical to the endogenous albumin molecules in blood, they are fully capable of utilizing the physiological albumin pathways. Indeed, nab-paclitaxel uses transcytosis mediated by albumin, enhancing intracellular paclitaxel delivery, and should take advantage of the overexpression of albumin-binding proteins such as Secreted Protein Acid and Rich Cysteine (SPARC) in stroma fibroblasts surrounding the tumor tissue. This should increase the selective uptake of this drug in tumor cells [17,18]. SPARC has indeed been evaluated as a biomarker of the activity for gemcitabine/nab-paclitaxel regimens because 1) it should enhance the selective uptake of this drug in tumor cells and 2) SPARC protein expression has been correlated to cancer cell proliferation and metastatic features [19]. However, *in vivo* experimental studies in both patient-derived xenografts and genetically engineered mouse models showed that SPARC did not play a role in nab-paclitaxel internalization. In addition, despite initial promising data, immunohistochemical analyses in PDAC specimens from patients treated with gemcitabine/nab-paclitaxel showed that SPARC levels were not associated with clinical outcome [20].

Similarly, we do not have biomarkers which can predict the toxicity of the gemcitabine/nab-paclitaxel regimen, and a recent multicentre retrospective observational study in the South Eastern Region of Sweden showed that neutropenia, leukopenia, thrombocytopenia, and anemia were observed in 23%, 20%, 5%, and 4% of patients treated with this regimen, respectively [21].

2.2. Folfirinox

FOLFIRINOX is a therapeutic regimen based on a combination of a number of chemotherapeutic drugs including 5-FU, irinotecan, leucovorin, and oxaliplatin. Following the positive results of the phase III clinical trial PRODIGE-4/ACCORD-11, this regimen has been introduced as a standard treatment in metastatic PDAC [22]. In an initial phase II trial, FOLFIRINOX was tested on 46 patients with advanced PDAC, showing a response rate of 26% (including 4% complete response), a median time to progression of 8.2 months and a median OS of 10.2 months [23]. Therefore, considering also the good safety profile, and improved in quality of life (QOL), FOLFIRINOX was further assessed in the above-mentioned phase III trial. In particular, since oncologists were lacking data on the efficacy and safety of FOLFIRINOX as compared with gemcitabine, the randomized controlled trial enrolled 342 patients with histologically and/or cytologically confirmed metastatic pancreatic adenocarcinoma and who had not previously been treated with chemotherapy. The median overall survival was 11.1 months in the FOLFIRINOX group compared to 6.8 months in the gemcitabine group ($P < 0.001$), while the median progression-free survival (PFS) was 6.4 months in the FOLFIRINOX and 3.3 months in the gemcitabine group ($P < 0.001$). Moreover, FOLFIRINOX regimen was associated with an objective response rate of 31.6%. Unfortunately, this regimen has also shown a severe grade 3–4 toxicity profile with 45.7% of neutropenia, 5.4% of febrile neutropenia, 9.1% of

thrombocytopenia, 14.5% of vomiting and 12.7% of diarrhea. In conclusion, as compared with gemcitabine, FOLFIRINOX was associated with a survival advantage but it had also increased toxicity [22].

In this regard, many efforts have been made in order to reduce the toxic effects. A prospective phase II open-label study evaluated a modified version of FOLFIRINOX (mFOLFIRINOX) which consisted in 25% dose reductions of irinotecan and bolus 5-FU given every 2 weeks (until progression, unacceptable toxicity, or surgical resection) [24]. mFOLFIRINOX given along with prophylactic pegfilgrastim was associated with a similar response rate and improved tolerability compared to full dose FOLFIRINOX [24].

Since the bolus 5-FU contributes to the toxicity, Mahaseth and colleagues proposed a modified FOLFIRINOX regimen which included discontinuation of the bolus 5-FU and administration of growth factors to all patients. Therefore, on day-3 after chemotherapy, prophylactic pegfilgrastim (6 mg) was administered subcutaneously to each patient. This modified FOLFIRINOX regimen showed a decreased incidence of grade 3–4 neutropenia to 3%, with a satisfactory response rate (30%), showing an improved safety profile with maintained efficacy in metastatic PDAC [25]. Similar results were obtained in patients treated with FOLFIRINOX at 80% dose intensity with routine use of growth factor support [26].

A phase III randomized trial of The Gruppo Oncologico Nord Ovest (GONO) further demonstrated the efficacy of the simplified FOLFOXIRI regimen in metastatic colorectal cancer. This modified regimen included a higher dose of 5-FU continuous infusion and a slightly lower dose of irinotecan [27]. More recently, Vivaldi and colleagues used the GONO-FOLFOXIRI regimen (irinotecan 165 mg/m^2 over 1 h, followed by oxaliplatin 85 mg/m^2 and leucovorin 200 mg/m^2 concomitantly over 2 h through a Y-connector, on Day-1, followed by fluorouracil 3,200 mg/m^2 as a 48-h continuous infusion starting on Day-1) and a modified schedule (irinotecan 150 mg/m^2 over 1 h, followed by oxaliplatin 85 mg/m^2 and leucovorin 200 mg/m^2 concomitantly over 2 h through a Y-connector, on Day-1, and followed by 5-FU 2,800 mg/m^2 as a 48-h continuous infusion starting on Day-1) in 137 stage III/IV PDAC patients. One (0.6%) complete response and 52 (38%) partial responses were observed in the whole population, with a disease control rate of 72.2%. The median OS and median PFS were 12 and 8 months, respectively. Regarding the toxicity profile, the main hematologic grade 3–4 toxicity was neutropenia (35.7%), but only one patient (0.7%) experienced febrile neutropenia. The main G3-4 non-hematological adverse events included G3 diarrhea in 11 (8%), nausea in 10 (7.3%), stomatitis in 9 (6.5%) and liver toxicity in 6 (4.4%) patients [28].

In conclusion, FOLFIRINOX introduction is arguably one of the most important innovations in PDAC care since the introduction of gemcitabine in 1996. This regimen has indeed been shown to dramatically improve OS, but this comes at the price of significantly increased toxicity (neutropenia, febrile neutropenia, thrombocytopenia, diarrhea, neuropathy), despite careful patient selection and the generous use of growth factor support. Moreover, clinical outcome (in terms of both efficacy and toxicity) in individual patients is relatively unpredictable, even when current clinical selection criteria (young age, good PS, absence of a biliary stent) are employed. Here we propose to describe and analyze new tools to improve the current usage of FOLFIRINOX as well as of gemcitabine/nab-paclitaxel, exploring new potential pharmacogenetic markers and treatments.

3. New tools to overcome PDAC resistance

3.1. Personalizing treatments using microRNA as predictive biomarkers of response

The prognosis of patients with PDAC is very poor because of the inherent and/or acquired resistance to conventional treatment modalities, which are also causing severe toxicities, as described above for the FOLFIRINOX and gemcitabine/nab-paclitaxel regimens [29]. Despite the constant efforts to formulate new chemotherapy regimens, new strategies to target these treatments are urgently warranted in order to achieve significant clinical improvement [30,31].

In this regard, the discovery of appropriate markers could help to determine which tumors will respond to which treatments, predicting the likelihood of drug resistance and leading treatment decisions.

In the last few years, microRNAs (miRNAs) have emerged as predictive biomarkers of response to conventional anti-cancer treatments and miRNAs-based strategies could represent a promising method to select the most appropriate pharmacological agent for personalization of the treatment [32].

Chemoresistance in PDAC is mediated by both genetic or epigenetic alterations and miRNAs play a key role in the epigenetic control. Indeed, miRNAs quickly respond to the genotoxic stress environment caused by chemotherapy and quickly modulate mRNA translation in cancer cells [4]. Most importantly, each miRNA controls the expression of multiple gene transcripts offering the possibility to identify critical miRNAs that could be used as informative biomarkers for detection, diagnosis, and prognosis of tumors that result from the deregulation of multiple genes. For this reason, a number of miRNA profiling studies have been conducted to obtain diagnostic and prognostic signatures for a variety of tumor type [33]. Therefore, several miRNAs playing a crucial role in the regulation of gene expression offer new directions for the quest of cancer biomarkers [34]

In the last few years, miRNAs expression in PDAC tissues has been largely studied and the PDAC miRNome has been extensively profiled [35]. However, only a few studies evaluated the role of candidate miRNAs to predict the sensitivity/resistance to conventional chemotherapy.

Several studies have shown the role of miR-21 expression levels in predicting which patients achieve the optimal response. For instance, Hwang and colleagues evaluated miR-21 expression levels in two cohorts of PDAC patients treated with gem and 5-FU and their results suggested that miR-21 expression can affect the outcome of both gemcitabine and 5-FU-based treatment [36]. The association of miR-21 expression levels and chemoresistance to gemcitabine was also observed in a study conducted on laser-microdissected specimens and PDAC cells. In particular, tissues isolated by laser microdissection were collected from 81 patients with

metastatic (n = 31) or nonmetastatic (n = 50) PDAC and normal ductal samples. The results of this study demonstrated a correlation between high expression of miR-21 and worse outcome after gemcitabine treatment. In particular, the authors reported a significant association between miR-21 expression and both disease-free survival and OS. Patients with high miR-21 expression had a significantly shorter OS both in the metastatic and in the adjuvant setting. Moreover, miR-21 expression in primary cultures correlated with expression in their respective tissues and with gemcitabine resistance. Further analyses on PDAC cells delineated the mechanism of action by which miR-21 induces gemcitabine resistance. It has been indeed observed that miR-21-downregulated PTEN and was associated with the activation of PI3K/Akt/mTOR pathway, reducing apoptosis induction by gemcitabine [37].

More recently, Wei and colleagues have proposed a similar mechanism underlying the role of miR-21 in resistance to 5-FU in human PDAC cells through the downregulation of tumor suppressor genes, including PTEN and PDCD4 [38]. Their findings confirmed high expression levels of miR-21 in resistant primary cells in comparison to sensitive primary cells and demonstrated that the suppression of miR-21 expression sensitized cancer cells to 5-FU treatment. On the contrary, the overexpression of mir-21 conferred resistance to 5-FU and promoted proliferation, migration, and invasion of PATU8988 and PANC-1 cells.

However, a phase III randomized trial evaluated the expression of miR-21 in tumor cells or cancer-associated fibroblasts (CAFs) in a cohort of 229 PDAC patients treated with 5-FU o gemcitabine. The expression levels of miR-21 were assessed by in situ hybridization, showing that miR-21 expression in CAFs was associated with decreased OS in PDAC patients who received 5-FU, but not gemcitabine. Conversely, strong expression of miR-21 in tumor cells was not correlated with survival in gemcitabine or 5-FU treated patients [39].

MiR-21 is not the only miRNA that seems to play a crucial role in chemoresistance to cytotoxic agents. In fact, Donahue and colleagues have revealed that miR-142-5p would be a promising predictive marker for gemcitabine treatment in patients with resected PDAC. First, they identified 24 miRNAs candidates that were up- or down-regulated in gemcitabine resistant cells. The analysis of miRNA expression, in relation to the survival time of PDAC patients after curative resection, revealed that miR-142-5p expression correlated with survival in patients treated with gemcitabine after surgical resection of PDAC, but not in patients without gemcitabine treatment. These findings could aid to select the appropriate and efficient treatment of patients after resection of PDAC, improving their prognosis [40].

More recently, *in vitro* and *in vivo* analyses have revealed a correlation between miR-506 expression levels and chemosensitivity to gemcitabine in PDAC cells. This study has highlighted that, after the transfection with miR-506 mimics, the overexpression of this miRNA enhanced the chemosensitivity of the cells to gemcitabine, whereas miR-506 inhibitors significantly conferred chemoresistance to PDAC cells. Moreover, sequencing analysis have revealed that the hypermethylation of the promoter region of the miR-506 gene reduced the expression levels of this miRNA in PDAC and significantly associated with poor prognosis. Despite these promising findings, further examination are needed to confirm the possibility of using miR-506 as a predictive biomarker of response to gemcitabine [41].

MiR-509-5p and miR-1243 have also been proposed as potential biomarkers in PDAC. A recent study showed that the overexpression of miR-509-5p and miR-1243 increased the sensitivity of PDAC cells to gemcitabine, suggesting that the expression status of these two miRNAs might predict gemcitabine efficacy in patients with PDAC [42].

Moreover, Preis and colleagues proposed miR-10b as an important tool for clinicians to guide clinical decision-making about neoadjuvant treatment and surgery. The authors of this study suggested, for the first time, that miR-10b expression in samples from fine needle aspirates (FNA) can be used to delineate a subgroup of patients that will truly benefit from subsequent surgery. In fact, patients whose cancers express low levels of miR-10b are predicted to have greater than 50% survival rate after 2 years. In contrast, high levels of miR-10b predict poor outcome and early disease progression even after surgical resection. Moreover, miR-10b expression levels in FNA samples correlated with the response to multimodality neoadjuvant gemcitabine-based chemoradiotherapy, in patients with resectable or locally advanced disease [43].

Of note, it has been demonstrated that also miR-211 expression modulates chemoresistance to gemcitabine. This miRNA emerged from a microarray analyses in long versus short-survival patients. *In vitro* studies showed that the overexpression of miR-211, by transfection of PDAC cells with pre-miR-211, led to a significant reduction of the percentages of cell growth and an increased sensitivity to gemcitabine. Instead, transfection with anti-miR-211 resulted in the opposite effect. Most likely, miR-211 modulates sensitivity to gemcitabine through the direct control of RRM2 expression. Indeed, the overexpression of miR-211 leads to a reduction of RMM2 expression levels, which represents a target of gemcitabine activity [44]. These results were further confirmed in a more recent study, where Maftouh and colleagues explored the biological role of miRNA-211 in gemcitabine activity in two subclones of SUIT2 cell line (SUIT2-028 and SUIT2-007). Their results revealed that the less aggressive subclone SUIT2-028, which was more sensitive to gemcitabine than the more aggressive subclone SUIT2-007, reported higher miR-211 expression levels [45].

MiR-200b, miR-200c and many members of the tumor suppressor let-7 family have been also suggested as a potential biomarker of gemcitabine resistance. Indeed, it has been found that their expression was significantly down-regulated in gemcitabine-resistant cells [46].

Other possible miRNAs that could represent good candidates as predictive biomarkers of response to chemotherapy are miR-192 and miR-215. Indeed, their overexpression results in reduced levels of thymidylate synthase which is the main drug target of the fluoropyrimidine/(5-FU)-based therapy [47].

Not many data are available on miRNAs affecting nab-paclitaxel, but several miRNAs have been associatedwith resistance to paclitaxel. For instance, it has been demonstrated that the ectopic expression of miR-200c downregulated

TUBB3 and enhanced sensitivity to microtubule-targeting agents, including paclitaxel [48]. Furthermore, miR-17-5p has been identified as one of most significantly downregulated miRNAs in paclitaxel-resistant lung cancer cells and it could affect the sensitivity to paclitaxel through the regulation of beclin1, which is one of the most important autophagy modulators [49]. On the contrary, it has been observed that miR-17-5p is upregulated in PDAC where it is associated with a poor prognosis and plays important roles in pancreatic carcinogenesis and progression [50]. However, further studies are needed to better evaluate the role of this miRNA as a predictive biomarker in PDAC.

A recent study evaluated the use of circulating miRNAs to predict and/or monitor patients with advanced PDAC treated with FOLFIRINOX. Results obtained showed that a reduction of miR-181a-5p expression levels in plasma is able to monitor response to FOLFIRINOX chemotherapy in patients with advanced PDAC. Conversely the same miRNA did not predict the outcome of patients treated with gemcitabine/nab-paclitaxel. *In vitro* analysis confirmed these findings and allowed to study the potential mechanism exploited by miR-181a-5b to affects FOLFIRINOX sensitivity. Interestingly, the inhibition of miR-181a-5p led to increased expression of tumor-suppressor protein ATM, enhancing sensitivity to oxaliplatin. However, the exact mechanisms of miR-181a-5p in chemosensitivity to FOLFIRINOX therapy remain to be investigated and the clinical utility of miR-181a-5p as a predictive biomarker should be further validated in prospective, large-scale clinical studies [51].

3.2. Targeting cancer cell glycolytic metabolism

3.2.1. Metabolic reprogramming in primary tumor and cancer metastasis

Pioneer Otto Warburg first revealed that metabolic differences exist between malignant tumor cells and adjacent normal cells. The neoplastic diseases are indeed characterized by a chronic and often uncontrolled cell proliferation which results in corresponding adjustments of energy metabolism in order to fuel cell growth and division. Despite the presence of oxygen, cancer cells switch from oxidative phosphorylation (OXPHOS) to the aerobic glycolysis, resulting in high rate glycolysis followed by lactic acid fermentation. Therefore, cancer cells tend to promote glycolysis over mitochondrial respiration, even under aerobic conditions. In tribute to Otto Warburg, this metabolic alteration is known as 'Warburg effect' [52,53].

Though aerobic glycolysis is an inefficient way to generate energy, cancer cells use this peculiar metabolic adaptation to increase the uptake and the incorporation of nutrients (such as nucleotides, amino acids, and lipids) into the biomass, conferring to cancer cell the advantage to obtain sooner all the elements they need to produce new cells [54].

Of note, the Warburg effect is a metabolic phenomenon involving the primary tumor but once cancer cell begin to spread from the original tumor to other organs or tissue of the body, their energy requirements change [55]. Emerging data suggest that metabolic flexibility is required for the success of the metastatic dissemination and is critical for efficient colonization of distant sites. Cancer cells differentially engage distinct metabolic strategies depending on their metastatic site [56]. For instance, breast cancer cells enhance their metastatic fitness to the liver by engaging a dominant metabolic phenotype. Instead, lung and bone metastasis are more dependent on OXPHOS [57].

3.2.2. Anti-cancer agents targeting the Warburg effect

Anti-cancer agents targeting the Warburg Effect include small inhibitory molecules that target glycolysis through the inhibition of glucose uptake (mainly mediated by Glucose transporter 1 (GLUT-1)), glucose retention (due to the reaction catalyzed by hexokinase (HK)) and lactate production (through the inhibition of LDH-A) (Figure 1). Glycolysis initiation requires glucose uptake that is mediated by the glucose transporters GLUTs [58], and overexpression of GLUT-1 has been found in various tumor types. This evidence can be explained by the high glucose consumption rate of cancer cell, which can be compensated by increased glucose influx through the overexpression of this transporter [59].

Of note, Shibuya and colleagues supposed that the increased glycolytic metabolism, characteristically seen in human cancers, has a pivotal role in the maintenance of cancer stem cells (CSCs) in several human cancer types. *In vitro* and *in vivo* analysis demonstrated that WZB117, a specific GLUT-1 inhibitor, could inhibit the self-renewal and tumor-initiating capacity of the human CSCs from different cancer types, including PDAC. Therefore, glucose metabolism-targeted therapy through GLUT-1 inhibition, in combination with other therapeutic modalities that effectively control the tumor bulk, is expected to achieve better management of cancers through a CSC-directed cancer therapy [60]. Moreover, a recent study conducted on primary PDAC resistant cells demonstrated that the novel GLUT-1 inhibitor PGL13 supports the synergistic interaction between the Akt inhibitor perifosine and gemcitabine, restoring the repression of aerobic glycolysis induced by Akt inhibitors [61].

Once glucose is inside the cell, the first critical and irreversible step of glycolysis is performed by HK, which coverts glucose to glucose-6-phosphate to ensure that glucose will not diffuse back out of the cells [62,63]. For this reason, HK is considered another promising potential target for antiglycolytic therapies. As regard to the use of HK inhibitors in PDAC, Bhardway and colleagues demonstrated that 3-bromopyruvate (3BP), an inhibitor of HK-II, used in combination with IAA, an inhibitor of GAPDHase, effectively inhibited both energy production and cell signaling in the PDAC cell line PANC-1 [64].

Moreover, a recent study performed on PDAC cell lines, showed that human *in vitro* derived tumor-associated macrophages (TAMs) utilize Warburg metabolism to promote tumor growth and dissemination. Indeed, this alteration in macrophage metabolism supported angiogenesis, and augmented the extravasation of tumor cells in a VEGF-dependent manner. Furthermore, it induced epithelial-to-mesenchymal transition (EMT) in a TGFβ-dependent manner, facilitating tumor dissemination to secondary sites (such as the lung and liver). Instead, the inhibition of glycolysis in TAMs, with a competitive inhibitor to HK2, 2-deoxyglucose (2DG), was

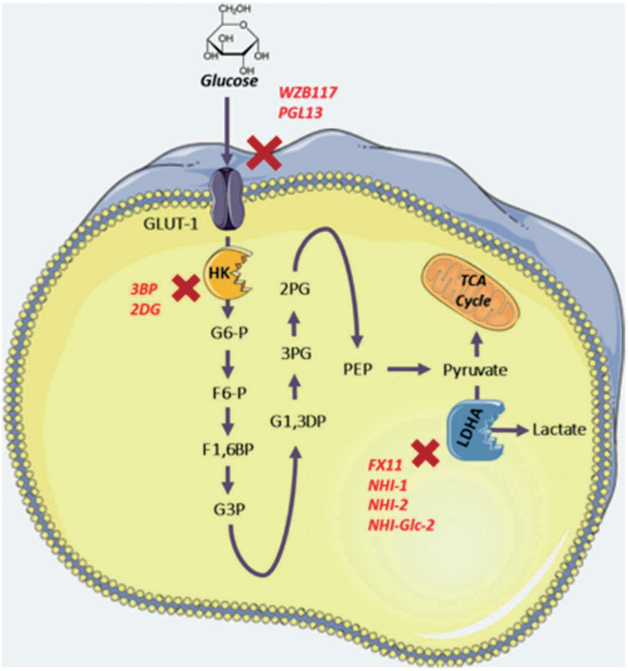


Figure 1. Scheme of selected components of the glycolysis pathway and inhibitors by small molecules targeting glucose transporter 1 (GLUT-1), hexokinase (HK) and lactate dehydrogenase (LDH-A) that have been studied or are undergoing evaluation as new anticancer treatments.

sufficient to disrupt this pro-metastatic phenotype, reversing the observed increases in TAM-supported angiogenesis, extravasation, and EMT [65]. In a phase I dose-escalation trial, the combination of 2-DG and docetaxel was tested in 34 patients with metastatic or advanced solid tumors. The results obtained prompt that the recommended dose of 2DG in combination with weekly docetaxel is 63 mg/kg/day with tolerable adverse effects [66], and further studies in PDAC patients seem warranted.

In the last step of glycolysis, pyruvate and ATP are formed. Then, pyruvate is taken up into mitochondria, converted into acetyl coenzyme A (Acetyl-CoA) which enters the tricarboxylic acid (TCA) cycle. But pyruvate can also be reduced to lactate by LDH-A, which has been proven to be overexpressed in many types of solid cancer, including PDAC [67]. FX11, a small-molecule inhibitor of LDH-A, negatively affects cellular energy supply, decreases cellular production of lactate, induces oxidative stress and finally provokes cell death. Of note, FX11 efficiently reduced cancer cell growth in PDAC cell and xenograft models [68,69]. This drug has been also tested in a more aggressive human pancreatic tumor xenograft LZ10.7, where it has been proven to be effective even as a single agent [68]. Unfortunately, to date, no clinical trial has been conducted to evaluate the clinical activity of FX-11.

A more recently developed class of effective inhibitors of LDH-A include N-hydroxyindole-based compounds [70]. Maftouh and colleagues demonstrated the synergistic interaction between gemcitabine and two derivatives of these compounds (NHI-1 and NHI-2). Indeed, they revealed that in hypoxic models of PDAC, these two compounds affected apoptosis, spheroid-growth, and invasiveness, reducing the expression of metalloproteinases. The synergistic interaction with gemcitabine was attributed to modulation of gemcitabine metabolism, overcoming the reduced synthesis of phosphorylated metabolites in hypoxia [71].

The glucose-conjugated methyl ester NHI-Glc-2 is a weaker LDH-A inhibitor than NHI-2 on the isolate enzyme. But anyway, it exploits the GLUT-1 overexpression, leading to an increased uptake [72]. The combination of NHI-Glc-2 with deoxynyboquinone (DNQ) has been already tested in cancer cells and in a mouse model of lung cancer [73]. Moreover, preliminary *in vitro* and *in vivo* studies showed a potential synergistic interaction between NHI-Glc-2 and gemcitabine in PDAC models [74].

Oncogenic *KRAS* mutations play a pivotal role in the initiation and progression of pancreatic adenocarcinoma. These mutations prevent GAP stimulation and lead to an accumulation of persistently GTP-bound and active K-Ras. In its activated form, K-Ras can trigger multiple signaling pathways, such as MAPK and P13, impacting a wide range of cellular processes, including cell survival and proliferation [75]. To date, mutant K-Ras remains difficult to target and some progress has been made only regarding the selective inhibition of G12C mutant *KRAS* isoform with a small molecule, which is more common in lung cancer compared to PDAC [76,77].

However, several recent studies have reported that *KRAS* plays a critical role in controlling cancer metabolism through

the stimulation of glucose metabolism, differential channeling of glucose intermediates, reprogrammed glutamine metabolism, increased autophagy, and micropinocytosis [78–81]. These findings could provide a novel and appealing alternative opportunity to targeting such a key determinant in PDAC. In fact, specific compounds designed to target the effector pathways of KRAS, can both alter PDAC cells metabolism and impair the ability of the cancer cells to proliferate.

Notably, Ying and colleagues, using inducible *KRAS*-G12D-driven PDAC mouse model, have already established that mutated *KRAS* enhances the expression of GLUT1 and several rate-limiting glycolytic enzymes, including HK and LDH-A [82]. Moreover, preclinical studies conducted on HK2-knockout models have shown that the inhibition of this glycolytic enzyme may provide therapeutic benefit to mutant K-Ras lung tumors [83]. Similarly, Xie and colleagues have reported that inactivation of LDH-A, in mouse models of NSCLC driven by oncogenic *KRAS* or *EGFR*, leads to decreased tumourigenesis and disease regression in established tumors [84].

In conclusion, metabolic targeting strategies that have been preclinically validated in lung cancer models could also represent a valid method to effectively target K-Ras in other tumor types. In particular, further studies on the use of anti-cancer agents targeting the Warburg effect in pancreatic tumors driven by *KRAS* are warranted.

4. Conclusions

In the last few years, several progresses have been made in chemotherapeutic treatment of PDAC, but current available treatments provide a limited survival advantage and are associated with severe toxicities. During the last decade, the growing interest in glycolytic metabolism of cancer cells has generated hope that a new class of effective anti-cancer treatment strategies could finally be developed. In fact, over the years, numerous types of small molecules have been tested for their hypothesized ability to hampers glucose uptake and glycolysis, representing a promising tool to target the Warburg effect in cancer cells. Unfortunately, only a small portion of these small molecules has been tested in clinical trials and novel preclinical and clinical development strategies are warranted.

Personalized medicine could overcome, at least in part, the current limits in chemotherapeutic treatment of PDAC, having the potential to tailor therapy with the best response and highest safety margin to ensure better patient care. In this regard, miRNAs have been increasingly recognized as promising predictive biomarkers of response to chemotherapeutic agents. In fact, according to a number of studies, miR-21, miR-181a and other miRNAs seems to predict response to specific treatments in PDAC cells and/or patients. However, further studies are needed to confirm their possible use in clinical practice as well as to find new candidate miRNAs to predict response to new therapeutic strategies.

5. Expert opinion

Alterations of miRNAs expression have an impact on patient-specific pharmacokinetics and pharmacodynamics of cancer drugs, rendering them closely related with resistance or sensitization to specific cytotoxic drugs. For this reason, analysis of miRNAs expression could represent a valuable tool to predict the response to currently available treatments and to help tailor treatment appropriate for individual patients [85].

MiRNA expression can be evaluated in tissues specimens as well as in a tissue biopsy of the primary tumor. In fact, in clinical practice, the source of biological material typically comes from formalin-fixed paraffin-embedded tumor samples obtained during standard of care surgical procedures or biopsies. Unfortunately, the tissue extraction is risky and painful for the patient and in some cases the sample is limited because of 'inaccessible' tumor localization. Moreover, the procedure to obtain a tissue biopsy is expensive, and cannot be applied repeatedly [86]. In addition, because of intratumour genetic heterogeneity, the molecular profile of the primary tumor from the initial surgical specimen might significantly differ from the molecular profile in a tumor sample obtained from a biopsy and might not reflect molecular aberrations accumulated as a consequence of selection pressure caused by cancer therapies [87]. Importantly, liquid biopsies overcome most of the limitations of tissue biopsies as they represent a noninvasive, rapid, precise method, which can also bypass the heterogeneity of tumors. Therefore, analysis of miRNAs expression in liquid biopsies, such as in serum or plasma or other body fluids, could represent a more effective way to predict the response to cancer drugs and hopefully would replace tissue biopsy in the near future [88].

In this review, we provided an overview of miRNAs that might predict response to standard treatments in PDAC. However, we also underlined the importance of new treatments, focussing on therapies targeting aberrant cancer glycolytic metabolism. These two topics should be further studied. To date, only a few studies evaluated miRNAs that could predict response to drugs targeting cancer cell glycolytic metabolism.

In particular, further studies should be conducted on miRNAs that, through their altered expression, might affect the expression of GLUT-1, HK2 or LDH-A (Figure 2). For instance, it has been demonstrated that miR-138, miR-150, miR-199a-3p and miR-532-5p downregulate GLUT-1 expression, whereas miR-19a, miR-19b, miR-130b, and miR-301a increase GLUT-1 expression in renal cell carcinoma. In addition, miR-144, which is downregulated in PDAC tissues and PANC-1 cells [89], was found to target GLUT-1 in lung cancer [90]. Therefore, the altered expression of these microRNAs, through the regulation of GLUT-1 expression, could modulate the chemoresistance toward GLUT-1 inhibitors such as WZB117 and PGL13.

In breast cancer cells miR-143 negatively regulate HK2 expression, whereas miR-155 indirectly promote HK2 transcription by repressing miR-143 [91]. Therefore, further studies on these two microRNAs could confirm their possible use as a marker of response to HK2 inhibitors such as *3BP* and *2DG*.

Finally, several microRNAs including miR-34a, miR-34c, miR-369-3p, miR-374a, miR-30a-5p, miR-142-3p, miR-30d-5p, miR-323a-3p, miR-199a-3p, miR-449a, and miR-4524a/b, have been found to target the mRNA of LDH-A [92–101]. In particular, it has been demonstrated that miR-34a re-sensitizes colorectal cancer cells to 5-FU by directly targeting the expression of LDH-A. These results suggest further studies on this miRNA as a predictive factor for response to a combined therapy of 5-FU with LDH-A inhibitors, such as FX11, NHI-1, NHI-2 and NHI-Glc-2 [102].

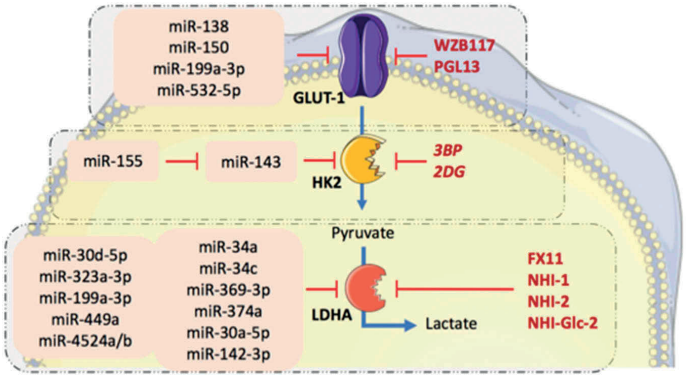


Figure 2. Scheme of potential mechanisms underlying the interaction or predictive potential of selected miRNAs with inhibitors of cancer glycolytic pathway. Four miRNAs (miR-138, miR-150, miR-199a-3p, miR-532-5p) can inhibit the glucose transporter GLUT-1, potentially modulating the sensitivity of cancer cells to the WZB117 and PGL13 compounds. Similarly, inhibition of HK2 can be modulated by a combinatorial effect of miR-143 and 3BP / 2DG inhibitors, while mir-155 can be used as predictive marker for 3BP or 2DG treatment response. A wide list of miRNAs have been found to target LDH-A and they may synergize the inhibitory effect on LDHA with FX11, NHI-1, NHI-2, NHI-Glc-2 small-molecule inhibitors.

In conclusion, future studies should evaluate the role of several emerging miRNAs as biomarkers in PDAC as well as to validate their use as potential predictive factors for the response to glycolysis inhibitors. Moreover, additional candidate miRNAs could be found by investigating the metabolism of glycolysis inhibitors. Indeed, miRNAs could modulate the expression of several drug-metabolizing enzymes, affecting the sensitivity to this class of new drugs. Hopefully, the identification of biomarkers of response, including the above-mentioned miRNAs, should shorten the gap between preclinical studies and personalized therapies using these novel treatments.

Funding

Funding from Italian Association for Italian Research (AIRC Start-Up grant) and Fondazione Pisana per la Scienza grant.

Declaration of interest

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

Reviewer disclosures

Peer reviewers on this manuscript have no relevant financial or other relationships to disclose.

References

Papers of special note have been highlighted as either of interest (•) or of considerable interest (••) to readers.

1. Rawla P, Sunkara T, Gaduputi V. Epidemiology of pancreatic cancer: global trends, etiology and risk factors. World J Oncol. 2019;10:10–27.
2. Kleeff J, Korc M, Apte M, et al. Pancreatic cancer. Nat Rev Dis Primers. 2016;2:16022.
3. Paulson AS, Tran Cao HS, Tempero MA, et al. Therapeutic advances in pancreatic cancer. Gastroenterology. 2013;144:1316–1326.
4. Guo S, Fesler A, Wang H, et al. microRNA based prognostic biomarkers in pancreatic cancer. Biomark Res. 2018;6:18.
5. Grasso C, Jansen G, Giovannetti E. Drug resistance in pancreatic cancer: impact of altered energy metabolism. Crit Rev Oncol Hematol. 2017;114:139–152.
6. Wu W, Zhao S. Metabolic changes in cancer: beyond the Warburg effect. Acta Biochim Biophys Sin (Shanghai). 2013;45(1):18–26.
7. Ciccolini J, Serdjebi C, Peters GJ, et al. Pharmacokinetics and pharmacogenetics of Gemcitabine as a mainstay in adult and pediatric oncology: an EORTC-PAMM perspective. Cancer Chemother Pharmacol. 2016;78:1–12.
8. Burris HA, Moore MJ, Andersen J, et al. Improvements in survival and clinical benefit with gemcitabine as first- line therapy for patients with advanced pancreas cancer: a randomized trial. J Clin Oncol. 1997;15:2403–2413.
9. Cunningham D, Chau I, Stocken DD, et al. Phase III randomized comparison of gemcitabine versus gemcitabine plus capecitabine in patients with advanced pancreatic cancer. J Clin Oncol. 2009;27:5513–5518.
10. Heinemann V, Quietzsch D, Gieseler F, et al. Randomized phase III trial of gemcitabine plus cisplatin compared with gemcitabine alone in advanced pancreatic cancer. J Clin Oncol. 2006;24:3946–3952.
11. Moore MJ, Goldstein D, Hamm J, et al. Erlotinib plus gemcitabine compared with gemcitabine alone in patients with advanced pancreatic cancer: a phase III trial of the National Cancer Institute of Canada Clinical Trials Group. J Clin Oncol. 2007;25:1960–1966.
12. Santarpia M, Rolfo C, Peters GJ, et al. On the pharmacogenetics of non-small cell lung cancer treatment. Expert Opin Drug Metab Toxicol. 2016;12:307–317.
13. Propper D, Davidenko I, Bridgewater J, et al. Phase II, randomized, biomarker identification trial (MARK) for erlotinib in patients with advanced pancreatic carcinoma. Ann Oncol. 2014;25:1384–1390.
14. Van Cutsem E, Li CP, Nowara E, et al. Dose escalation to rash for erlotinib plus gemcitabine for metastatic pancreatic cancer: the phase II RACHEL study. Br J Cancer. 2014;111:2067–2075.
15. Von Hoff DD, Ervin T, Arena FP, et al. Increased survival in pancreatic cancer with nab-paclitaxel plus gemcitabine. N Engl J Med. 2013;369:1691–1703.
16. Cucinotto I, Fiorillo L, Gualtieri S, et al. Nanoparticle albumin bound paclitaxel in the treatment of human cancer: nanodelivery reaches prime-time? J Drug Deliv. 2013;2013:1–10.

17. Desai N, Trieu V, Yao Z, et al. Increased antitumor activity, intratumor paclitaxel concentrations, and endothelial cell transport of cremophor-free, albumin-bound paclitaxel, ABI-007, compared with cremophor-based paclitaxel. Clin Cancer Res. 2006;12:1317–1324.
18. Desai N. Drug delivery report winter. Technol Overviews. 2008:37–41.
19. Neuzillet C, Tijeras-Raballand A, Cros J, et al. Stromal expression of SPARC in pancreatic adenocarcinoma. Cancer Metastasis Rev. 2013;32:585–602.
20. Hidalgo M, Plaza C, Musteanu M, et al. SPARC expression did not predict efficacy of nab-paclitaxel plus gemcitabine or gemcitabine alone for metastatic pancreatic cancer in an exploratory analysis of the phase III MPACT trial. Clin Cancer Res. 2015;21:4811–4818.
21. Blomstrand H, Scheibling U, Bratthäll C, et al. Real world evidence on gemcitabine and nab-paclitaxel combination chemotherapy in advanced pancreatic cancer. BMC Cancer. 2019;19:40.
22. Conroy T, Desseigne F, Ychou M, et al. FOLFIRINOX versus gemcitabine for metastatic pancreatic cancer. N Engl J Med. 2011;364:1817–1825.
23. Conroy T, Paillot B, François E, et al. Irinotecan plus oxaliplatin and leucovorin-modulated fluorouracil in advanced pancreatic cancer - a groupe tumeurs digestives of the Fédération Nationale des Centres de Lutte Contre le Cancer study. J Clin Oncol. 2005;23:1228–1236.
24. James ES, Yao X, Cong X, et al. Interim analysis of a phase II study of dose-modified FOLFIRINOX (mFOLFIRINOX) in locally advanced (LAPC) and metastatic pancreatic cancer (MPC). J Clin Oncol. 2014;32:256–256.
25. Mahaseth H, Brutcher E, Kauh J, et al. Modified FOLFIRINOX regimen with improved safety and maintained efficacy in pancreatic adenocarcinoma. Pancreas. 2013;42:1311–1315.
26. Lowery MA, Yu KH, Adel NG, et al. Activity of front-line FOLFIRINOX (FFX) in stage III/IV pancreatic adenocarcinoma (PC) at Memorial Sloan-Kettering Cancer Center (MSKCC). J Clin Oncol. 2012;30:4057.
27. Falcone A, Ricci S, Brunetti I, et al. Phase III trial of infusional fluorouracil, leucovorin, oxaliplatin, and irinotecan (FOLFOXIRI) compared with infusional fluorouracil, leucovorin, and irinotecan (FOLFIRI) as first-line treatment for metastatic colorectal cancer: the gruppo oncologico nor. J Clin Oncol. 2007;25:1670–1676.
28. Vivaldi C, Caparello C, Musettini G, et al. First-line treatment with FOLFOXIRI for advanced pancreatic cancer in clinical practice: patients' outcome and analysis of prognostic factors. Int J Cancer. 2016;139:938–945.
29. Collisson EA, Olive KP. Pancreatic cancer: progress and challenges in a rapidly moving field. Cancer Res. 2017;77:1060–1062.
30. Garrido-Laguna I, Hidalgo M. Pancreatic cancer: from state-of-the-art treatments to promising novel therapies. Nat Rev Clin Oncol. 2015;12:319–334.
31. Neoptolemos JP, Kleeff J, Michl P, et al. Therapeutic developments in pancreatic cancer: current and future perspectives. Nat Rev Gastroenterol Hepatol. 2018;15:333–348.
32. Chatterjee SK, Zetter BR. Cancer biomarkers: knowing the present and predicting the future. Futur Oncol. 2005;1:37–50.
33. Calin GA, Croce CM. MicroRNA signatures in human cancers. Nat Rev Cancer. 2006;6:857–866.
34. Esquela-Kerscher A, Slack FJ. Oncomirs - microRNAs with a role in cancer. Nat Rev Cancer. 2006;6:259–269.
35. Bloomston M, Frankel WL, Petrocca F, et al. MicroRNA expression patterns to differentiate pancreatic adenocarcinoma from normal pancreas and chronic pancreatitis. J Am Med Assoc. 2007;297:1901–1908.
36. Hwang JH, Voortman J, Giovannetti E, et al. Identification of microRNA-21 as a biomarker for chemoresistance and clinical outcome following adjuvant therapy in resectable pancreatic cancer Hoheisel J, editor. PLoS One. 2010;5:e10630.
37. Giovannetti E, Funel N, Peters GJ, et al. MicroRNA-21 in pancreatic cancer: correlation with clinical outcome and pharmacologic aspects underlying its role in the modulation of gemcitabine activity. Cancer Res. 2010;70:4528–4538.
• **Extensivestudy on the prognostic role of microRNA-21 in pancreatic cancer.**
38. Wei X, Wang W, Wang L, et al. MicroRNA-21 induces 5-fluorouracil resistance in human pancreatic cancer cells by regulating PTEN and PDCD4. Cancer Med. 2016;5:693–702.
39. Donahue TR, Nguyen AH, Moughan J, et al. Stromal microRNA-21 levels predict response to 5-fluorouracil in patients with pancreatic cancer. J Surg Oncol. 2014;110:952–959.
40. Ohuchida K, Mizumoto K, Kayashima T, et al. MicroRNA expression as a predictive marker for gemcitabine response after surgical resection of pancreatic cancer. Ann Surg Oncol. 2011;18:2381–2387.
41. Li J, Wu H, Li W, et al. Downregulated miR-506 expression facilitates pancreatic cancer progression and chemoresistance via SPHK1/Akt/NF-κB signaling. Oncogene. 2016;35:5501–5514.
42. Hiramoto H, Muramatsu T, Ichikawa D, et al. MiR-509-5p and miR-1243 increase the sensitivity to gemcitabine by inhibiting epithelial-mesenchymal transition in pancreatic cancer. Sci Rep. 2017;7:4002.
43. Preis M, Gardner TB, Gordon SR, et al. MicroRNA-10b expression correlates with response to neoadjuvant therapy and survival in pancreatic ductal adenocarcinoma. Clin Cancer Res. 2011;17:5812–5821.
44. Giovannetti E, van der Velde A, Funel N, et al. High-Throughput MicroRNA (miRNAs) arrays unravel the prognostic role of MiR-211 in pancreatic cancer Ellis NA, editor. PLoS One. 2012;7:e49145.
45. Maftouh M, Avan A, Funel N, et al. MiR-211 modulates gemcitabine activity through downregulation of ribonucleotide reductase and inhibits the invasive behavior of pancreatic cancer cells. Nucleosides Nucleotides Nucleic Acids. 2014;33:384–393.
46. Li Y, Vandenboom TG, Kong D, et al. Up-regulation of miR-200 and let-7 by natural agents leads to the reversal of epithelial-to-mesenchymal transition in gemcitabine-resistant pancreatic cancer cells. Cancer Res. 2009;69:6704–6712.
47. Boni V, Bitarte N, Cristobal I, et al. miR-192/miR-215 influence 5-fluorouracil resistance through cell cycle-mediated mechanisms complementary to its post-transcriptional thymidylate synthase regulation. Mol Cancer Ther. 2010;9:2265–2275.
48. Cochrane DR, Spoelstra NS, Howe EN, et al. MicroRNA-200c mitigates invasiveness and restores sensitivity to microtubule-targeting chemotherapeutic agents. Mol Cancer Ther. 2009;8:1055–1066.
49. Yan HJ, Liu WS, Sun WH, et al. MiR-17-5p inhibitor enhances chemosensitivity to gemcitabine via upregulating Bim expression in pancreatic cancer cells. Dig Dis Sci. 2012;57:3160–3167.
50. Yu J, Ohuchida K, Mizumoto K, et al. MicroRNA miR-17-5p is overexpressed in pancreatic cancer, associated with a poor prognosis and involved in cancer cell proliferation and invasion. Cancer Biol Ther. 2010;10:748–757.
51. Meijer LL, Garajová I, Caparello C, et al. Plasma miR-181a-5p downregulation predicts response and improved survival after FOLFIRINOX in pancreatic ductal adenocarcinoma. Ann Surg. 2018;1.
•• **Article of considerable interest because it is the first study suggesting a role for the modulation of a plasma microRNA in the prediction of response to FOLFIRINOX.**
52. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144:646–674.
53. Warburg O. The metabolism of carcinoma cells 1. J Cancer Res. 1925;9:148–163.
54. Heiden Vander MG, Cantley LC, Thompson CB. Understanding the warburg effect: the metabolic requirements of cell proliferation. Science. 2009;324:1029–1033.
55. Weber GF. Time and circumstances: cancer cell metabolism at various stages of disease progression. Front Oncol. 2016;6:257.
56. Lehúede C, Dupuy F, Rabinovitch R, et al. Metabolic plasticity as a determinant of tumor growth and metastasis. Cancer Res. 2016;76:5201–5208.
• **Seminal review on the role of metabolism in cancer progression and metastasis.**
57. Dupuy F, Tabariès S, Andrzejewski S, et al. PDK1-dependent metabolic reprogramming dictates metastatic potential in breast cancer. Cell Metab. 2015;22:577–589.
58. Navale AM, Paranjape AN. Glucose transporters: physiological and pathological roles. Biophys Rev. 2016;8:5–9.

59. Wang J, Ye C, Chen C, et al. Glucose transporter GLUT1 expression and clinical outcome in solid tumors: a systematic review and meta-analysis. Oncotarget. 2017;8:16875–16886.
60. Shibuya K, Okada M, Suzuki S, et al. Targeting the facilitative glucose transporter GLUT1 inhibits the self-renewal and tumor-initiating capacity of cancer stem cells [Internet]. Oncotarget. 2015;6:651–661.
• **Manuscript describing a new strategy for targeting GLUT1.**
61. Massihnia D, Avan A, Funel N, et al. Phospho-Akt overexpression is prognostic and can be used to tailor the synergistic interaction of Akt inhibitors with gemcitabine in pancreatic cancer. J Hematol Oncol. 2017;10:9.
62. Liberti MV, Locasale JW. The Warburg Effect: how does it benefit cancer cells? Trends Biochem Sci. 2016;41:211–218.
63. Akram M. Mini-review on glycolysis and cancer. J Cancer Educ. 2013;28:454–457.
64. Bhardwaj V, Rizvi N, Lai MB, et al. Glycolytic enzyme inhibitors affect pancreatic cancer survival by modulating its signaling and energetics. Anticancer Res. 2010;30:743–749.
65. Penny HL, Sieow JL, Adriani G, et al. Warburg metabolism in tumor-conditioned macrophages promotes metastasis in human pancreatic ductal adenocarcinoma. Oncoimmunology. 2016;5(8):e1191731.
• **Interesting study on the role of Warburg metabolism in tumor-conditioned macrophages, promoting metastasis in human pancreatic ductal adenocarcinoma.**
66. Raez LE, Papadopoulos K, Ricart AD, et al. A phase i dose-escalation trial of 2-deoxy-d-glucose alone or combined with docetaxel in patients with advanced solid tumors. Cancer Chemother Pharmacol. 2013;71:523–530.
67. Petrelli F, Cabiddu M, Coinu A, et al. Prognostic role of lactate dehydrogenase in solid tumors: a systematic review and meta-analysis of 76 studies. Acta Oncol (Madr). 2015;54:961–970.
68. Le A, Cooper CR, Gouw AM, et al. Inhibition of lactate dehydrogenase A induces oxidative stress and inhibits tumor progression. Proc Natl Acad Sci U S A. 2010;107:2037–2042.
69. Rajeshkumar NV, Dutta P, Yabuuchi S, et al. Therapeutic targeting of the warburg effect in pancreatic cancer relies on an absence of p53 function. Cancer Res. 2015;75:3355–3364.
70. Rani R, Granchi C. Bioactive heterocycles containing endocyclic N-hydroxy groups. Eur J Med Chem. 2015;97:505–524.
71. Maftouh M, Avan A, Sciarrillo R, et al. Synergistic interaction of novel lactate dehydrogenase inhibitors with gemcitabine against pancreatic cancer cells in hypoxia. Br J Cancer. 2014;110:172–182.
72. Calvaresi EC, Granchi C, Tuccinardi T, et al. Dual targeting of the warburg effect with a glucose-conjugated lactate dehydrogenase inhibitor. ChemBioChem. 2013;14:2263–2267.
73. Lee HY, Parkinson EI, Granchi C, et al. Reactive oxygen species synergize to potently and selectively induce cancer cell death. ACS Chem Biol. 2017;12:1416–1424.
74. El Hassouni B, Sciarrillo R, Mantini G, et al. PO-042 targeting hypoxic pancreatic cancer cells with glucose conjugated lactate dehydrogenase inhibitor nhi-glc-2. ESMO. 2018;3(Suppl 2):3082.
•• **Recent study describing a new approach to target LDH-A in pancreatic cancer models.**
75. Di Magliano MP, Logsdon CD. Roles for KRAS in pancreatic tumor development and progression. Gastroenterology. 2013;144:1220–1229.
76. Pant S, Hubbard J, Martinelli E, et al. Clinical update on K-Ras targeted therapy in gastrointestinal cancers. Crit Rev Oncol Hematol. 2018;130:78–91.
77. Lito P, Solomon M, Li L-S, et al. Allele-specific inhibitors inactivate mutant KRAS G12C by a trapping mechanism. Science. 2014;351:604–608.
78. Bryant KL, Mancias JD, Kimmelman AC, et al. Feeding pancreatic cancer proliferation [Internet]. Trends Biochem Sci. 2014;39:91–100.
79. Cohen R, Neuzillet C, Tijeras-Raballand A, et al. Targeting cancer cell metabolism in pancreatic adenocarcinoma. Oncotarget. 2015;6:16832–16847.
80. Kimmelman AC. Metabolic dependencies in RAS-driven cancers. Clin Cancer Res. 2015;21:1828–1834.
• **Important piece on metabolic aberrations in *KRAS*-Driven Cancers.**
81. White E. Exploiting the bad eating habits of Ras-driven cancers. Genes Dev. 2013;27:2065–2071. Cold Spring Harbor Laboratory Press.
82. Ying H, Kimmelman AC, Lyssiotis CA, et al. Oncogenic kras maintains pancreatic tumors through regulation of anabolic glucose metabolism. Cell. 2012;149:656–670.
83. Patra KC, Wang Q, Bhaskar PT, et al. Hexokinase 2 is required for tumor initiation and maintenance and its systemic deletion is therapeutic in mouse models of cancer. Cancer Cell. 2013;24:213–228.
84. Xie H, Hanai JI, Ren JG, et al. Targeting lactate dehydrogenase-A inhibits tumorigenesis and tumor progression in mouse models of lung cancer and impacts tumor-initiating cells. Cell Metab. 2014;19:795–809.
85. Detassis S, Grasso M, Del Vescovo V, et al. microRNAs make the call in cancer personalized medicine. Front Cell Dev Biol. 2017;5.
86. Marrugo-Ramírez J, Mir M, Samitier J. Blood-based cancer biomarkers in liquid biopsy: a promising non-invasive alternative to tissue biopsy. Multidisciplinary Digital Publishing Institute (MDPI) [Internet]. Int J Mol Sci. 2018;19:1.
87. Janku F. Tumor heterogeneity in the clinic: is it a real problem? [Internet]. Ther Adv Med Oncol. 2014;6:43–51. SAGE Publications.
88. Ponti G, Manfredini M, Tomasi A. Non-blood sources of cell-free DNA for cancer molecular profiling in clinical pathology and oncology. Crit Rev Oncol Hematol. 2019;141:36–42.
• **Recentreview on potential non-blood sources for studies on liquid biopsies.**
89. Lan F, Yu H, Hu M, et al. MiR-144-3p exerts anti-tumor effects in glioblastoma by targeting c-Met. J Neurochem. 2015;135:274–286.
90. Liu M, Gao J, Huang Q, et al. Downregulating microRNA-144 mediates a metabolic shift in lung cancer cells by regulating GLUT1 expression. Oncol Lett. 2016;11:3772–3776.
91. Jin LH, Wei C. Role of microRNAs in the Warburg effect and mitochondrial metabolism in cancer. Asian Pacific J Cancer Prev. 2014;15:7015–7019. Asian Pacific Organization for Cancer Prevention.
92. Wang J, Wang H, Liu A, et al. Lactate dehydrogenase A negatively regulated by miRNAs promotes aerobic glycolysis and is increased in colorectal cancer. Oncotarget. 2015;6:19456–19468.
93. Xiao X, Huang X, Ye F, et al. The MIR-34a-LDHA axis regulates glucose metabolism and tumor growth in breast cancer. Sci Rep. 2016;6:21735.
94. Ping W, Senyan H, Li G, et al. Increased lactate in gastric cancer tumor-infiltrating lymphocytes is related to impaired T cell function due to miR-34a deregulated lactate dehydrogenase A. Cell Physiol Biochem. 2018;49:828–836.
95. Li X, Lu P, Li B, et al. Sensitization of hepatocellular carcinoma cells to irradiation by MIR-34a through targeting lactate dehydrogenase-A. Mol Med Rep. 2016;13:3661–3667.
96. Li L, Kang L, Zhao W, et al. miR-30a-5p suppresses breast tumor growth and metastasis through inhibition of LDHA-mediated Warburg effect. Cancer Lett. 2017;400:89–98.
97. Hua S, Liu C, Liu L, et al. miR-142-3p inhibits aerobic glycolysis and cell proliferation in hepatocellular carcinoma via targeting LDHA. Biochem Biophys Res Commun. 2018;496:947–954.
98. He Y, Chen X, Yu Y, et al. LDHA is a direct target of miR-30d-5p and contributes to aggressive progression of gallbladder carcinoma. Mol Carcinog. 2018;57:772–783.
99. Chen H, Gao S, Cheng C. MiR-323a-3p suppressed the glycolysis of osteosarcoma via targeting LDHA. Hum Cell. 2018;31:300–309.
100. Zhou S, Min Z, Sun K, et al. miR-199a-3p/Sp1/LDHA axis controls aerobic glycolysis in testicular tumor cells. Int J Mol Med. 2018;42:1786–1798.
101. Li L, Liu H, Du L, et al. MiR-449a suppresses LDHA-mediated glycolysis to enhance the sensitivity of non-small cell lung cancer cells to ionizing radiation. Oncol Res. 2018;26:547–556.
102. Li X, Zhao H, Zhou X, et al. Inhibition of lactate dehydrogenase A by microRNA.34a resensitizes colon cancer cells to 5.fluorouracil. Mol Med Rep. 2015;11:577–582.

Chapter 6

Blood platelet RNA yields a new diagnostic prospective for accurate detection and staging of pancreatic ductal adenocarcinoma

Meijer LL*, In 't Veld SGJG*, Le Large TYS*, Mantini G, van der Lelij E, Capula M, Dwarshuis G, Sol N, Liu D, Di Franco G, Post E, Verschueren H, Zonderhuis BM, Daams F, Besselink M, Bijlsma MF, Morelli L, Best MG, Frampton AE, Giovannetti E**, Wurdinger T**, Kazemier G**
* These authors contributed equally ** These authors contributed equally

Submitted

Chapter 7

Omics Analysis of Educated Platelets in Cancer and Benign Disease of the Pancreas

Mantini G, Meijer LL*, Glogovitis I*, In 't Veld SGJG, Paleckyte R, Morelli L, Le Large TY, Capula M, Pham T, Piersma S, Jimenez CR, Kazemier G, Koppers-Lalic D, Wurdinger T**, Giovannetti E**.
* These authors contributed equally ** These authors contributed equally

Cancers, 2021

cancers

MDPI

Article

Omics Analysis of Educated Platelets in Cancer and Benign Disease of the Pancreas

Giulia Mantini [1,2], Laura L. Meijer [1,3,†], Ilias Glogovitis [4,5,†], Sjors G. J. G. In 't Veld [4], Rosita Paleckyte [1], Mjriam Capula [2,6], Tessa Y. S. Le Large [1,3], Luca Morelli [7], Thang V. Pham [1], Sander R. Piersma [1], Adam E. Frampton [8,9], Connie R. Jimenez [1], Geert Kazemier [3], Danijela Koppers-Lalic [4], Thomas Wurdinger [4,*] and Elisa Giovannetti [1,2,*]

1 Department of Medical Oncology, Cancer Center Amsterdam, Amsterdam UMC, VU University Medical Center (VUmc), 1081 HV Amsterdam, The Netherlands; g.mantini@amsterdamumc.nl (G.M.); l.meijer@amsterdamumc.nl (L.L.M.); rosita.paleckyte@polpharmabiologics.com (R.P.); t.lelarge@amsterdamumc.nl (T.Y.S.L.L.); t.pham@amsterdamumc.nl (T.V.P.); s.piersma@amsterdamumc.nl (S.R.P.); c.jimenez@amsterdamumc.nl (C.R.J.)
2 Fondazione Pisana per la Scienza, 56017 Pisa, Italy; m.capula@fpscience.it
3 Department of Surgery, Cancer Center Amsterdam, Amsterdam UMC, VU University Medical Center (VUmc), 1081 HV Amsterdam, The Netherlands; g.kazemier@amsterdamumc.nl
4 Department of Neurosurgery, Cancer Center Amsterdam, Amsterdam UMC, VU University Medical Center (VUmc), 1081 HV Amsterdam, The Netherlands; ilias@uni-plovdiv.bg (I.G.); g.intveld1@amsterdamumc.nl (S.G.J.G.I.V.); d.lalic@amsterdamumc.nl (D.K.-L.)
5 Department of Plant Physiology and Molecular Biology, University of Plovdiv, 4002 Plovdiv, Bulgaria
6 Institute of Life Sciences, Sant'Anna School of Advanced Studies, 56127 Pisa, Italy
7 General Surgery Unit, Department of Translational Research and New Technologies in Medicine and Surgery, University of Pisa, 56126 Pisa, Italy; luca.morelli@unipi.it
8 Department of Clinical and Experimental Medicine, Faculty of Health and Medical Sciences, The Leggett Building, University of Surrey, Guildford GU2 7WG, UK; a.frampton@imperial.ac.uk
9 Faculty of Health and Medical Sciences, The Leggett Building, University of Surrey, Guildford GU2 7XH, UK
* Correspondence: t.wurdinger@amsterdamumc.nl (T.W.); e.giovannetti@amsterdamumc.nl (E.G.); Tel.: +31-003-120-444-2633 (E.G.)
† These authors contributed equally to this paper.

Simple Summary: Tumor cells are known to produce and secrete pro-coagulants that recruit blood particles such as platelets, inducing hypercoagulability. However, platelets can also influence tumor carcinogenesis and metastasis, creating a reciprocal, vicious loop with the tumors. Confrontation of platelets with tumor cells via transfer of tumor-associated biomolecules or influencing platelets biology ("education") is an emerging concept, that has been recently proposed to create innovative platforms for biomarkers within blood-based "liquid biopsies". In this study, we explore the intrinsic regulation and the potential "education" of platelets using -omics profiling in pancreatic cancer patients. Our results showed: (i) a high activity on RNA splicing that can lead to subsequent platelets education; (ii) enrichment of specific modified forms (isomiRs) of canonical miRNAs; and (iii) inhibition of SPARC transcription by specific class of isomiRs. Moreover, we created an interactive tool to visualize expected correlations, to facilitate further investigations on additional potential biomarkers and therapeutic tools.

Abstract: Pancreatic ductal adenocarcinoma (PDAC) is traditionally associated with thrombocytosis/hypercoagulation and novel insights on platelet-PDAC "dangerous liaisons" are warranted. Here we performed an integrative omics study investigating the biological processes of mRNAs and expressed miRNAs, as well as proteins in PDAC blood platelets, using benign disease as a reference for inflammatory noise. Gene ontology mining revealed enrichment of RNA splicing, mRNA processing and translation initiation in miRNAs and proteins but depletion in RNA transcripts. Remarkably, correlation analyses revealed a negative regulation on SPARC transcription by isomiRs involved in cancer signaling, suggesting a specific "education" in PDAC platelets. Platelets of benign

Citation: Mantini, G.; Meijer, L.L.; Glogovitis, I.; In 't Veld, S.G.J.G.; Paleckyte, R.; Capula, M.; Le Large, T.Y.S.; Morelli, L.; Pham, T.V.; Piersma, S.R.; et al. Omics Analysis of Educated Platelets in Cancer and Benign Disease of the Pancreas. *Cancers* **2021**, *13*, 66. https://doi.org/10.3390/cancers13010066

Received: 17 November 2020
Accepted: 24 December 2020
Published: 29 December 2020

Cancers **2021**, *13*, 66

patients were enriched for non-templated additions of G nucleotides (#ntaG) miRNAs, while PDAC presented length variation on 3′ (lv3p) as the most frequent modification on miRNAs. Additionally, we provided an actionable repertoire of PDAC and benign platelet-ome to be exploited for future studies. In conclusion, our data show that platelets change their biological repertoire in patients with PDAC, through dysregulation of miRNAs and splicing factors, supporting the presence of *de novo* protein machinery that can "educate" the platelet. These novel findings could be further exploited for innovative liquid biopsies platforms as well as possible therapeutic targets.

Keywords: liquid biopsy; platelets; omics integration; regulatory mechanisms; gene expression; miRNAs; proteins; pancreatic cancer

1. Introduction

Cancer death predictions in the US show pancreatic ductal adenocarcinoma (PDAC) expected to become the second leading cause of cancer-related deaths by 2030 [1]. The majority of PDAC patients are diagnosed late, with either locally-advanced or metastatic disease, which are typically resistant to chemotherapy. The lack of reliable biomarkers for preventive screening or early cancer detection, and the absence of effective therapies, are the main causes for the poor survival rates, ranging between 2 and 9% [2,3].

Tumor cells are known to secrete pro-coagulants or fibrinolytic substances that recruit platelets and this can induce hypercoagulability [4,5]. This is particularly common in patients suffering from PDAC, with an incidence of thrombotic complications up to 36%, and can be attributed to the procoagulant properties of PDAC cells, including the promotion of platelet activation [6–8].

Recent studies reported the use of platelets as an extremely promising source for cancer diagnosis and early detection biomarkers [9–13]. Platelets do not have a nucleus but hold a pool of megakaryocyte-derived mRNAs [14,15] and the complete machinery for *de novo* protein synthesis, resulting in dynamic modifications of protein expression [16]. Remarkably, platelets contain a vast amount of bioactive proteins which can be secreted upon activation [17]. These proteins can be either synthesized by the platelets themselves or taken up during circulation, making platelets profiling an extremely appealing tool to obtain a representative "image" of the current status of the healthy or diseased body.

It has already been shown that platelets can alter their RNA profiles when cancer cells are present, and they are referred as "tumor educated platelets" (TEPs) [18]. On this basis, previous studies generated robust classifiers to identify the cancer status based on the platelet signature in different tumor types, including PDAC [19–21]. However, further studies to investigate the molecular mechanisms underlying the "education" of platelets are warranted.

Evidence that platelets are capable of de novo protein synthesis [22], raised the issue of whether there is a fine-tuning of their content depending on external stimuli. During RNA splicing, intronic sequences of pre-mRNA are generally removed, while exonic sequences are joined together. However, the splicing process can create several mRNA sequences by varying the pre-RNA composition (e.g., by retaining introns or skipping exons), within a process called alternative splicing. Denis et al. identified pre-mRNA splicing as a displaced nuclear process that can occur in platelets [23]. Moreover, the presence of retained introns transcripts has been suggested to be a biologically relevant phenomenon that contribute to modulation of the platelet transcriptome [16,24].

Apart from RNA, circulating platelets are also enriched in small non-coding RNAs (ncRNAs) such as miRNAs [25–27]. It is well recognized that many small ncRNAs play a pivotal role in regulation of mRNAs expression in physiological and pathological conditions [28–30]. In particular, miRNAs bind to specific regions in the target mRNAs, thus leading to mRNA degradation or repression, subsequently resulting in suppression of protein translation [31–34].

Most miRNA genes are transcribed by RNA polymerase II in the nucleus. The pri-miRNA is then cleaved by the microprocessor Drosha. The resulting pre-miRNA is then exported by Exportin-5 in the cytoplasm where it is further cleaved by Dicer. Commonly, one strand of the miRNA is bound to AGO protein and incorporated in the RISC complex that will repress gene transcription [35]. However, malfunctions of Drosha and Dicer can create modified forms of miRNA, called "isomiRs". Additionally, isomiRs can be generated by the addition of nucleotides to the 3′ or 5′ ends by nucleotidyltransferases such as TUT4 and GLD2. Remarkably, there are several types of isomiRs, depending on the miRNA length and the addition of nucleotides with respect to the canonical form, and they have already been described as potential biomarkers for prostate cancer detection [29].

A successful integration of data from those types of molecules (i.e., miRNAs, isomiRs, mRNA and protein expression) can lead to the discovery of new biomarkers that reflect their complex relationships as well as to understand which biological pathways are affected in different diseases, including cancer [36].

Here, to further unravel the biology underlying TEP profiles in PDAC, we applied for the first-time parallel deep omics approaches using Next Generation Sequencing (NGS) for RNA and small-RNA and label-free LC-MS/MS for proteomics, using benign pancreatic disease as "control group" to elucidate the biology of platelets in patients affected by malignant tumor platelets.

These studies provide an extensive catalog of -omics data that can be used in many ways to explore non-coding RNAs, mRNAs and protein expression in platelets of PDAC and benign lesions. To facilitate further research on diagnostic markers from non-invasive biopsies, novel targets to inhibit metastases formation and the many further uses that can be envisaged, all data are available in GEO (GSE160252), proteomeXchange (PXD022514) and resulting biological networks from these analyses are presented in the R Shiny Web App: http://platelnet.eu.ngrok.io while the R script code is available at: https://github.com/Giulia221091/Platel-net.

2. Results

First, we investigated whether the inclusion of healthy donors can be useful in our study setting. We analyzed age- and sex- matched healthy donors samples (HD, n = 19) and their RNA expression profile when compared to PDAC and benign platelets. Based on the different RNA profile of HD (Figure S1) we have excluded this group from the following analyses. Indeed, HD cases classify aside from the other groups, suggesting that the expression of RNA from blood platelets of healthy individuals is different from the RNA expression of platelets from patients with PDAC or benign diseases. This difference might be explained by the fact that all patients with benign diseases had some inflammatory responses which could result in differences in RNA expression in platelets compared to HD.

Our choice to compare PDAC to benign disease was also sustained by the actual clinical challenge: distinguish PDAC patients from patients having non-malignant disease. Unfortunately, clinical symptoms and diagnostic features of patients with PDAC show a considerable similarity to those of patients with different benign diseases of the pancreas. Currently, the diagnostic process relies on clinical suspicion, radiological investigation, brush cytology or fine-needle aspiration for pathological confirmation, and measurement of tumor markers. However, most clinically used biomarkers fail to discriminate PDAC from benign diseases, substantiating the need for correction of the inflammatory signal in -omic analyses.

To obtain a comprehensive overview of the regulation of the transcriptome and proteome in platelets of PDAC and benign disease, we performed quantitative proteomic analysis (~2000 identified proteins), small-RNA profiling (~44,000 canonical and isomiRs type identified) and transcriptomics analysis (~50,000 RNA transcripts identified) on isolated and highly purified platelets. Platelets were collected from patients with PDAC (N = 11) and age- and sex matched patients with benign disease (N = 11). Clinical charac-

teristics and age- and sex distribution are presented in Table S1 and Figure S2. In addition, proteomics analysis was performed on a subset with sufficient protein content after carefully checking that age and sex were still matching between groups in proteomics and transcriptomics datasets (Figure S2).

Figure 1 describes the workflow overview adopted in this study. Blood samples were collected from patients with PDAC and benign disease, with subsequent platelets isolation. After measuring and processing for small-RNAs, transcriptomics and label-free mass spectrometry-based proteomics, we evaluated differential expressed profiles for each data type. Moreover, we performed an intra group correlation analysis in PDAC and benign patients and an integrative gene ontology mining to understand the common enriched pathways between data types in PDAC and benign platelets.

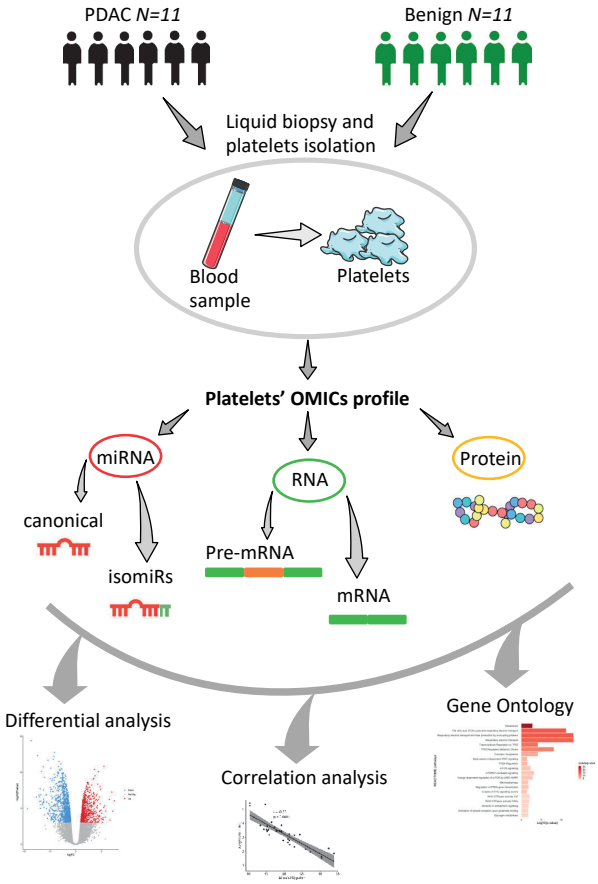


Figure 1. Workflow of platelets omics profiling. Platelets were isolated from 22 age and sex-matched patients with PDAC and benign diseases. Small-RNAs, RNA transcripts (PDAC samples = 11; benign samples = 11) and proteins (PDAC samples = 8; benign samples = 11) were isolated and sequenced following validated protocols [28,37]. Bioinformatics tools were adopted to quantify canonical as well as isomiRs from smallRNA-seq, mRNA and intron-spanning reads from RNA-seq, and proteins. Downstream analyses were carried out using standalone tools for differential analysis of all data types comparing PDAC versus benign platelets; intra group correlation analysis between miRNAs, mRNAs and proteins of matched PDAC and benign platelets and, lastly, gene ontology mining.

A repertoire of 344 canonical miRNAs, 8357 isomiRs, 8695 intron-spanning reads, 49965 mRNAs and 2106 proteins were used for downstream analyses as reported in Figure 2A. Next, differential expression analysis was computed separately for each data-type including canonical miRNAs, isomiRs, intron-spanning reads, mRNAs and proteins (Figure 2B,C).

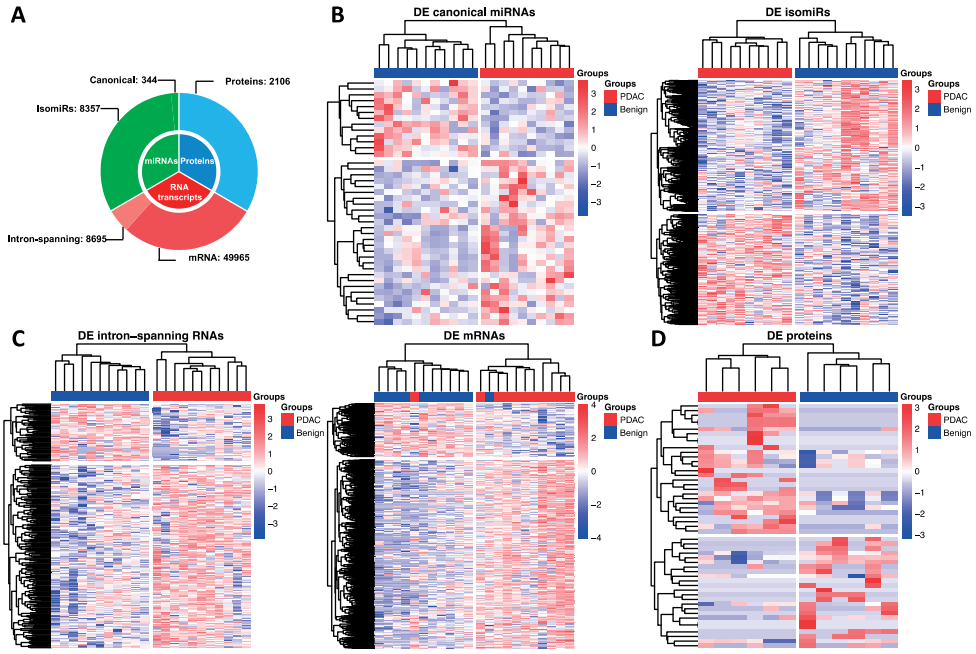


Figure 2. Data summary and differential analysis. (**A**) summary of canonical miRNAs, isomiRs, mRNAs, intron-spanning reads and proteins used in this study for subsequent analysis; (**B**) heatmaps of differentially expressed canonical miRNAs and isomiRs; (**C**) heatmaps of differentially expressed intron-spanning reads and mRNAs; (**D**) heatmap of differentially expressed proteins. Significance level was set to p-value < 0.05 and $|\text{Log}_2\ \text{FC}| > 1$.

2.1. Differential Analysis

First, we aimed to differentiate platelets of PDAC and benign patients using each data-type. This analysis led to the identification of: (i) 41 differentially expressed (DE) canonical miRNAs (28 up-regulated in PDAC and 13 up-regulated in benign platelets) (Figure 2B left panel); (ii) 981 DE isomiRs (448 up-regulated in PDAC and 533 up-regulated in benign) (Figure 2B right panel); (iii) 285 DE intron-spanning reads (217 up-regulated in PDAC and 68 up-regulated in benign) (Figure 2C left panel); (iv) 1878 DE mRNAs (1466 up-regulated in PDAC and 412 up-regulated in benign) (Figure 2C right panel); (v) 52 DE proteins (26 up-regulated in PDAC and 26 in benign) (Figure 2D). Raw-sequencing data can be found at www.ncbi.nlm.nih.gov under the accession GSE160252.

Regarding the canonical miRNAs up-regulated in PDAC platelets, we observed that many of them were expected to regulate genes of ECM-receptor interaction. In particular, miR-128, miR-29a and miR-335 were the miRNAs targeting more genes involved in this interaction. However, the same miRNAs were targeting genes enriched for proteoglycans in cancer and PI3K-Akt signaling pathways. Those pathways have been described in several studies reporting their aberration in PDAC and other cancer types [38–41].

Analyzing the differentially expressed intro-spanning RNA reads we found MAP2K4, CDC42, CBL, SOS1, ROCK2, FOXO1 up-regulated in PDAC and associated to MAPK

signaling, insulin signaling, TGF-beta and PDGF pathways. Previous studies already showed the alteration of those signaling pathways in PDAC cells [42,43].

However, the analysis on mature-mRNA revealed enrichment of very broad terms such as cell adhesion, cytoskeletal protein binding, and ion binding. This suggests that activated platelets in PDAC patients may require several proteins to affect different processes. Indeed, we identified many proteins up-regulated in platelets from PDAC patients that are involved in poly-A RNA binding, mRNA binding and transferase activity, suggesting a clear "protein machinery" in action [16,44].

2.2. Regulatory Networks

Understanding the internal regulation of platelet activity in PDAC patients is one of the main goals for this study. Thus, we used the negative correlations between miRNAs (canonical and isomiRs) and genes (intron-spanning RNAs and mRNAs), and positive correlations between genes and proteins that we call "expected" correlations. This has led to 8 resulting networks in total: four networks for PDAC platelets and four networks for benign platelets. These four networks were based on expected correlations of: (i) isomiRs-mRNA-proteins; (ii) isomiRs-intronRNA-proteins; (iii) canonical miRNA–mRNA-proteins; (iv) canonical miRNA-intronRNA-proteins.

Analyzing the resulting four correlation networks from PDAC platelets we discovered that five RNA transcripts namely SNTB1, SPARC, PPM1A, TLN1 and ADD3 were represented in all four networks. Of those transcripts, SPARC showed to be between the top five most connected nodes with a node degree ranging from three (in canonical networks) to 32 (in isomiRs networks).

Specifically, the mRNA transcript of SPARC was found to be down-regulated in PDAC platelets patients. Figure 3 clearly show that SPARC down-regulation is mainly associated by isomiRs. Each connection represents one specific class of isomiR (e.g., nta#A, mv, nta#C etc.). The various number and type of isomiRs competing for SPARC down-regulation is given by the biological effects in which several miRNAs compete for the same gene target. In Figure 3A, we show the mRNA transcript SPARC is negatively correlated to several isomiRs, namely miR-17-3p, miR-29a-3p, miR-22-3p and miR-221-5p, while panel B shows canonical miRNAs associated to SPARC modulation that are not significantly different between PDAC and benign blood platelets included in our study. Of note, high expression of miR-22-3p is associated to poor survival in an external cohort of PDAC patients (Figure S3) as reported previously [45]. Moreover, KEGG pathways analysis of the above-mentioned miRNAs revealed an enrichment of classical PDAC pathways such as PI3K-Akt signaling, mTOR pathway, focal adhesion and "pancreatic cancer pathways" itself (Table S2). Additionally we have performed a multivariate analysis on SPARC and miR-29a-3p expression correcting for age, sex and stage and assessed that those clinical features are not confounding factors (Table S3).

The same analysis strategy was adopted in benign correlation networks. This time, none of the RNA transcripts was found to be over-represented between the four networks. However, we found spectrin-β non erythrocytic 1 (SPTBN1) of particular interest. This RNA transcript was one of the most connected nodes in isomiRs-networks and found to be down-regulated in benign platelets (Figure S4A,B) suggesting a key role of PDAC progression for this gene.

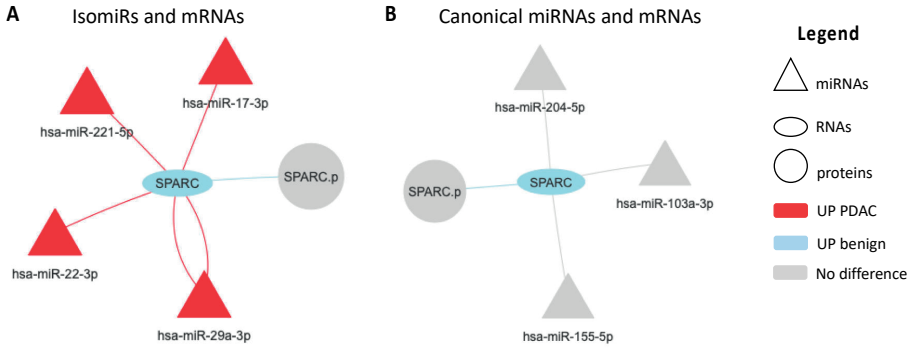


Figure 3. Networks focused on SPARC and based on correlation analyses of miRNAs, RNAs and proteins in PDAC platelets. A node-to-node connection is only seen if the nodes satisfy the expected correlation rule (negative correlation between miRNA and mRNAs, and positive correlation between mRNAs and proteins). (**A**) expected correlations between isomiRs and mRNA data; (**B**) expected correlations between canonical miRNAs and mRNA data; In isomiRs data each connection represents one isomiRs type, red and light blue colors represent up-regulation in PDAC and benign platelets, respectively. Triangle, oval and circle represents miRNAs, RNAs and proteins, respectively.

Previous analysis from TCGA consortium already showed a significant down-regulation of RNA transcript SPTBN1 in healthy and non-tumor matched samples when compared to PDAC tissues (Figure S4C). In addition, proteomics analysis confirmed transcriptomics results, showing elevated proteomic levels of SPTBN1 in $CD24^{+}$ PDAC tissues when compared with $CD24^{-}$ adjacent normal tissues [46]. However, divergent results on transcriptomics and proteomics data are found when studying prognosis of PDAC patients. Indeed, in a proteomics study of 55 pancreatic cancer patients, lower levels of SPTBN1 correlate with advanced PDAC stage and worse prognosis [47], and similar data were reported in a cohort of 82 resected PDAC patients [48] (Figure S3D). Contrary to this, transcriptomics data show that high levels of SPTBN1 correlates with poor prognosis. Therefore, different regulations at transcriptomics levels or post-translational modifications may tune SPTBN1 expression from benign state to cancer progression.

2.3. SPARC Is a Direct Target of miR-29a-3p and Its Modulation Affect Cell Migration

To validate SPARC as an important target of miR-29a-3p, we evaluated previous literature studies [49–53] using cloned 3′UTR regions of this transcript in luciferase vectors and co-expressed them with precursor miRNAs (pre-miRs). Most of these studies showed reduced luciferase levels upon miRNA over-expression and verified a direct miRNA-mRNA interaction (Figure 4A). We then transfected both Panc-1 and LPC006 cells with the pre-miRs and anti-miRs, individually, and performed RT-qPCR to assess and confirm changes at the endogenous mRNA levels. Transfection efficiency of pre- and anti-miR-29a-3p was evaluated by qRT-PCR analysis, 48 h post transfection, showing a significant modulation of miR-29a-3p expression in both cellular models (Figure S5). Consistent with the literature findings, we observed a reduction in the levels of SPARC mRNA in cells with increased expression of miR-29a-3p. On the contrary, we observed a significant increase of SPARC expression in cells with reduced miR-29a-3p expression (Figure 4B). The effect of miR-29a-3p on cell migration was evaluated using the wound healing assay, which showed that the hsa-miR-29a-3p mimic inhibited PDAC cell migration capability, while the hsa-miR-29a-3p inhibitor enhanced it (Figure 4C). These findings demonstrate that hsa-miR-29a-3p regulates the mRNA expression of its predicted target SPARC and affects PDAC cell migration.

A

Database	Experiment	Support type	Cell line	Cancer type	Reference
Mirecords	qRT-PCR	Functional MTI	HeLa and HepG2	Nasopharyngeal carcinomas	[49]
Mirtarbase	Luciferase reporter assay/Microarray/Northern blot/qRT-PCR/Western blot	Functional MTI	HeLa	Acute myeloid leukemia	[50]
Mirtarbase	qRT-PCR	Functional MTI	MCF-7	Breast cancer	[51]
Mirtarbase	Luciferase reporter assay/qRT-PCR/Western blot	Functional MTI	MHCC-97L and PLC	Hepatocellular carcinoma	[52]
Mirtarbase	Luciferase reporter assay	Functional MTI	IMR-90	Lung adenocarcinoma	[53]

MTI= miRNA-Target Interaction

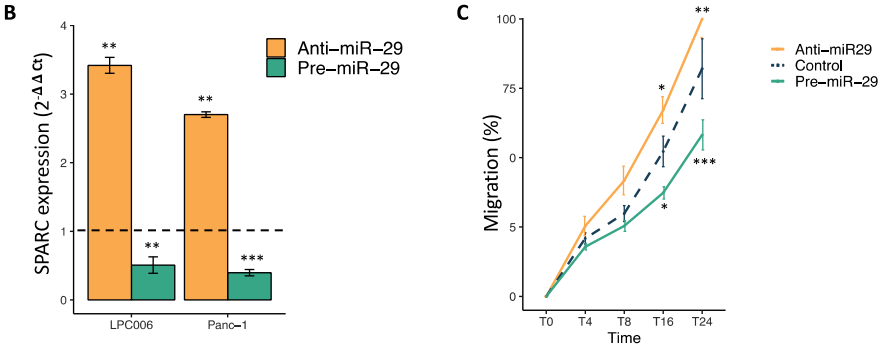


Figure 4. Validation of miR-29a-3p inhibitory effect on SPARC mRNA expression. (**A**) Table on previous studies validating the inhibition of SPARC caused by miR-29a-3p. (**B**) Modulation of SPARC mRNA expression levels after transfection with pre-miR-29a-3p (green) or anti-miR-29a-3p (orange) in LPC006 and Panc-1 cells. (**C**) Modulation of migration in LPC006 cells after transfection with anti-miR-29a-3p (orange), pre-miR-29a-3p (green) or miR-negative control (blue–dashed line). Significance was assessed with T-Student Test (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.0001$). Dashed line refers to comparative levels in miR-negative controls.

2.4. Integration of Gene Set Enrichment Analyses

Defining and discovering a regulatory mechanism in a specific phenotype is often challenging. For example, many miRNAs can compete to target the same gene and the same gene can be targeted by many miRNAs simultaneously. Not only miRNAs, but the entire family of small-RNAs can regulate gene expression by both mRNA degradation and translational repression mechanisms and it has been shown that protein expression not always correlate to gene expression due to post-translation modifications and/or regulatory feedbacks.

We hypothesized that the most active and prominent pathway that describes the current state of the cell/platelet should be maintained across all data-layers.

To this end, we performed a gene set enrichment analysis using Gene Ontology terms for each data-type. Next, results of the separate gene ontology mining were integrated retaining only the overlapped GO terms (Figure 5). Interestingly, a total of eight pathways resulted significantly enriched for RNA splicing, mRNA processing, ribosome biogenesis and translation initiation in miRNAs and proteins of PDAC platelets, in line with previous findings [16,44]. Genes and intron-spanning reads were found to be down-regulated in PDAC platelets. This can be well explained by the inhibitory mechanisms of miRNAs acting on gene expression. Of note, proteins that are supposed to correlate to RNA expression are on opposite direction. This outcome can have at least three explanations: (i) the presence of alternative regulatory mechanisms and post translational

modifications on RNAs; (ii) various proteins also non-tumor related can be ingested by platelets; (iii) blood platelets can ingest miRNAs and proteins that can control internal splicing events encoding for unfolded/non-functional proteins inducing a "specific" regulation. Indeed, down-regulation of RNA splicing in platelets of cancer patients was already shown by Best et al. [19] when biologically mining the selected features for their diagnostic TEP model.

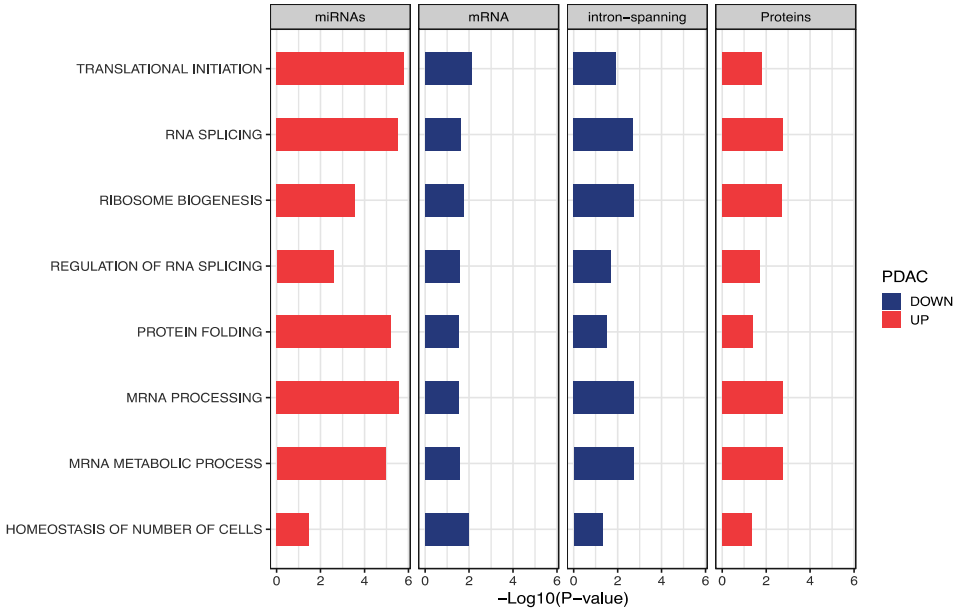


Figure 5. Gene ontology mining. Overlapped of significant GO terms for miRNAs, mRNAs, intron-spanning reads and proteins. Red bars describe biological terms enriched in PDAC platelets while blue bars identify biological terms down-regulated in PDAC platelets.

2.5. Different Isomirs Profile in PDAC and Benign Platelets of Patient

We performed an exploratory analysis on the different expression protein profile of eight PDAC and 11 benign platelets patients using our proteomics data and four additional datasets to validate our findings: (i) label-free quantification from DDA data of discovery cohort; (ii) label-free quantification from DIA data of discovery cohort; (iii) iq implementation of DIA data of discovery cohort [54]; (iv) published study on proteomics platelets using healthy controls and PDAC patients [9].

In this analysis, we found five highly expressed proteins in PDAC platelets: HBA1, HBD, PRDX2, CA1 and 1 down-regulated protein in PDAC platelets: AGT. Results are presented in Figure S6.

Next, we investigated whether there was a correlation with transcriptomics and small-RNAs data based on the above-mentioned analysis. Unfortunately, none of the six differentially expressed proteins were differentially regulated in transcriptomics data. However, AGT, HBD and CA1 shared a miRNA target that is up-regulated in PDAC, miR-26b-5p (Table S4).

Next, we generated an isomiRs profile based on differentially expressed canonical miRNAs (Figure 6). Based on this profile, miR-26b-5p is stable in benign platelets (only addition of #G base and the 5′ ends), while several types of isomiRs of this miRNA are produced in PDAC platelets (nta#A, lv5p, lv3p, mlv5p). We then investigated the frequency of each isomiR class in PDAC and benign platelets, using only the isoform of DE canonical

miRNAs (side barplots Figure 5). Strikingly, while benign platelets present a relatively high number of isomiRs of nta#G, in PDAC platelets this class of isomiR is almost absent and several different types of isomiRs appear to increase, with nta#T and lv3p being the most frequent. These results suggest a different action of nucleotidyltransferases in the two patient groups.

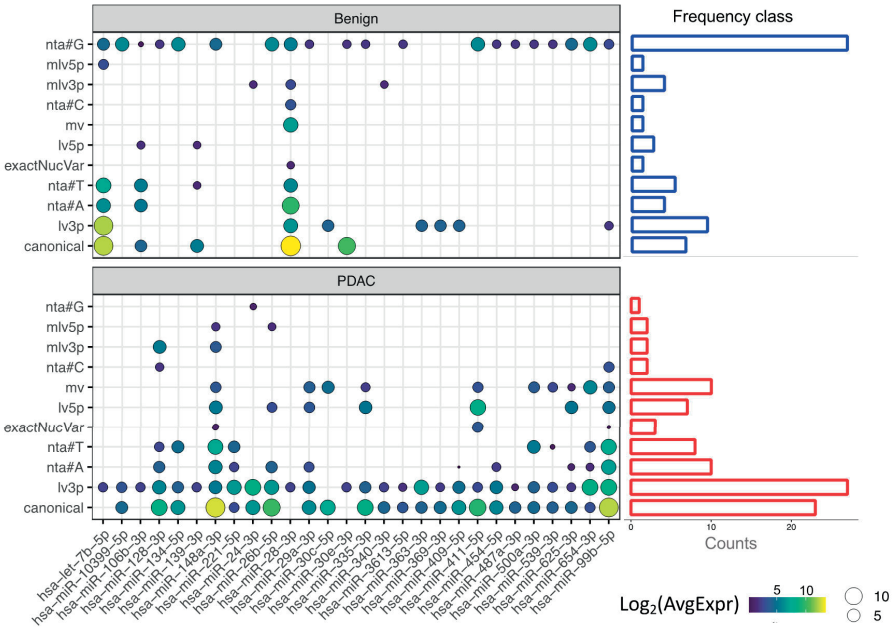


Figure 6. Heatmap of differentially expressed miRNAs and their average expression $\log_2$ transformed in benign (upper panel) and PDAC platelets (bottom panel) following the type of isomiR: (i) canonical; (ii) length variant in 3′; (iii) non-templated addition of base A; (iv) non-templated addition of base T; (v) exact nucleotide variant; (vi); length variant in 5′ (vii) multiple variant; (viii) non-templated addition of base C; (ix) multiple variant in 3′; (x) multiple variant in 5′; (xii) non-templated addition of base G. Side barplots represent the frequency of the isomiR class in benign platelets (blue) and PDAC platelets (red).

3. Discussion

This study demonstrates a novel integrative interaction of miRNAs, mRNAs, and protein expression in blood platelets of patients with PDAC, when compared to benign lesions. Parallel deep omics approaches were applied to unravel the biology of TEPs and correlate expression profiles. The intrinsic regulation of platelets was further explored using –omics profiling. With an extensive catalog of data profiles, we demonstrated profound regulations at multiple levels with high discriminatory power.

Next, correlation analysis of intron-spanning reads and mRNAs with corresponding proteins were compared in PDAC and benign platelets to evaluate the presence of different translational activity.

Importantly, SPARC resulted to be down-regulated in PDAC networks when using isomiRs data. In particular, SPARC is negatively correlated with several isomiRs, such as miR-221-5p, miR-29a-3p, miR-22-3p and miR-17-3p. We further validated the nhibitory effects of miR-29a-3p on SPARC expression through RT-qPCR in PDAC cells. Ontology mining for those miRNAs revealed an over-expression of classical PDAC pathways, such as ErbB signaling, PI3K-Akt signaling, mTOR pathway, focal adhesion. Notably, high plasma

levels of miR-22 appear to be prognostic for poor survival in an external PDAC cohort [45]. However, a clear relationship between miR-22 and SPARC has not yet been described.

SPARC is a multifunctional glycoprotein with different and somehow controversial activities, it can modulate cellular interaction with the extracellular matrix (ECM) by binding collagen and vitronectin, but also contributes to counteradhesive cells effects by focal adhesion abrogation. In tumorigenesis, SPARC presents a downregulated pattern in specific tumor cell types (e.g., epithelial) and an upregulated pattern in adjacent stromal cells, as described in ovarian, pancreatic and lung cancers [55–57]. SPARC negatively regulates cell proliferation, angiogenesis and adhesion, but is increased in gliomas (grades II–IV) [58]. These opposing actions of SPARC may be clarified by differences in the biological activities of several proteolytic molecules including matrix metalloproteinases, cathepsins, elastases and serine proteases [59]. Moreover, SPARC regulates the activity of several growth factors such as platelet-derived growth factor, basic fibroblast growth factor and vascular endothelial growth factor that can all play a pivotal role in platelet molecular mechanisms underlying cancer progression and metastases.

In this study, SPARC negatively correlates to the above-mentioned isomiRs suggesting that alterative mechanisms such as malfunctioning on proteins involved in miRNAs biogenesis or ingestion of tumor-secreted miRNAs can potentially play a role in regulation of SPARC gene transcripts and the subsequent inhibition of the tumor suppressor. However, a still open question is if SPARC is produced by platelets and released on the tumor site, or it is taken up by platelets from tumor cells.

Finally, a possible "pathway flow" from miRNAs to proteins going across mRNAs and retained introns mRNAs was investigated. Integration of significantly enriched gene ontology terms from differentially expressed miRNAs, mRNAs, intron-spanning reads and proteins showed that RNA splicing, RNA transcription, mRNA processing and translation initiation terms were enriched in PDAC platelets for miRNAs and proteins but not for mature mRNAs and intron-spanning reads. This result is in line with previous findings [60] suggesting that other mechanisms are acting at the level of mRNA processing. For example, we found that many miRNAs related to the RNA splicing pathway, where up-regulated in PDAC platelets, whereas the mature mRNAs associated to them where down-regulated. This may suggest that other regulatory mechanisms are involved in mature-RNA degradation such as 5′ and 3′ modifications on RNA extremities [61], RNA helicases, poly-A tail elongation, chaperones and silencing RNAs (psiRNAs) [62]. However, integrated pathway analysis showed that miRNAs correlated with protein levels, meaning that, also when mRNA is degraded, proteins are still produced. A potential explanation for such phenomenon could be that miRNAs and proteins secreted by CTCs or directly from the tumor are recruited and ingested by platelets.

We demonstrated that differentially expressed canonical miRNAs of PDAC blood platelets are enriched for different type of isomiRs class such as lv3p, nta#A and nta#T while benign platelets solely show an enrichment for nta#G. There are two possible explanation for these findings: (i) those miRNAs are not produced in the platelets but can be taken up after contact with tumor, CTCs or tumor vesicles secreted; (ii) platelets contain nucleotidyltransferases that can alter the miRNA template and the subsequent interaction with the gene target. Both mechanisms can result in a modification of the platelet transcriptome and its subsequent education in response to external stimuli such as the presence of the tumor.

A seminal study performed in 2018 developed a non-invasive blood test called CancerSEEK [63] that aimed to detect eight common cancer types based on eight circulating protein biomarkers and tumor-specific mutations in circulating DNA. Depending on the cancer type, this method detected cancer with a sensitivity ranging from 69 to 98% opening a promising future in this field. More recently, an increasing number of studies has shown that cell-free DNA (cfDNA) methylation could be utilized for the identification of disease-specific signatures in pre-neoplastic lesions or chronic pancreatitis (CP), representing a sensitive and non-invasive method of early diagnosis of PDAC. An exhaustive review by Gall and colleagues reported a summary of all cfDNA studies in PDAC, chronic

pancreatitis (CP) and benign lesions [64]. A major limiting factor of the reproducibility of (cf)DNA methylation data is the lack of a common reporting standard for DNA methylation detection. Furthermore, the heterogeneity of the cfDNA molecules and their genetic variability associated with cancer make the results' interpretation even more difficult. Despite technological advances in heterogeneity deconvolution [65–67] pinpointing of tissue- and disease-specific subsets of molecules remains a difficult task. More insights in tumor heterogeneity will surely boost the utilization of liquid biopsies and, thus, the discovery of PDAC biomarkers.

Of note, a recent analysis reported the first methylation landscape on tumor suppressor genes in PDAC [68]. In this study higher methylation indices for SPARC were able to distinguish PDAC from CP, confirming previous observations [69]. Furthermore, SPARC hyper-methylation was associated with stage IV, metastasized disease, and poor survival. However, this study was conducted in small cohorts, and validation in larger independent cohorts is warranted.

Overall, this is the first study where miRNAs, RNAs and proteins were profiled from the same set of platelets samples in PDAC patients. Having in mind the limitation of the sample size and the low detection rate for proteomics, this study aimed to explore the intrinsic regulation of platelets using -omics profiling. To this end, we generated an extensive catalog of data profiles and an interactive tool to visualize expected correlations, in order to facilitate further investigations on additional diagnostic biomarkers and therapeutic tools.

A recent study described how tumor educated platelets enable both brain tumor diagnostic and therapy monitoring [70]. The strategy adopted behind in this study interrogated intron-spanning RNA reads and their expression using a fine-tuned classifier (swarm intelligence). This resulted in an accuracy of 0.95% when discriminating glioblastoma from asymptomatic healthy controls. Remarkably, the same method was adopted to diagnose a non-tumor disease such as multiple sclerosis with 80% of accuracy [71]. These findings demonstrate that spliced-RNA of tumor educated platelets could be a used for different diseases.

To conclude, using a multi-omics profiling of platelets from PDAC and benign diseases patients we (i) provide additional data supporting previous findings where signature of RNA splicing is down-regulated in RNA of cancer platelets; (ii) illustrated that all the -omics profiles can classify the two groups; and (iii) showed that specific isomiRs types are present in PDAC platelets prompting future studies to further validate their biological and clinical relevance.

4. Materials and Methods

4.1. Patients and Samples Collection

This study enrolled 11 patients with pathologically confirmed PDAC and 11 patients with benign disease, both groups comprising male and female subjects aged between 39 and 81 years. Clinical characteristics of the study subjects are listed in Table S1.

Blood samples were obtained from the two University Hospitals of Amsterdam (VUmc and AMC, Amsterdam Universities Medical Centers (Amsterdam UMC, Amsterdam, The Netherlands), after receiving an informed consent of the patients (medical ethical approved protocol: #14438). Healthy donors were reported to be without any type of cancer, currently or in the past, as described previously [72]. The samples and associated clinical data of all individuals was collected and stored with a retraceable code, and fully anonymized.

4.2. Isolation of RNA and Protein for miRNAs, mRNAs and Protein Profiling

To elucidate potential different regulatory mechanisms, miRNA, mRNA and proteins were profiled from the same blood platelets samples.

Platelet pellets were isolated within 48 h after blood collection and stored in RNAlater at -80°. Total RNA was extracted from platelets using the MiRVAna kit (Ambion) while the protein fraction was stored for proteomics analysis after removal of RNA. Platelets isolation and extraction was performed as described previously [37].

4.2.1. RNA-Seq Library Preparation

For RNA isolation and sequencing, all samples were subjected to the thromboSeq protocol described previously [37].

4.2.2. miRNA-Seq Library Preparation

Preparation of miRNA libraries and sequencing was performed as described by Koppers-Lalic et al. [28,29]. After RNA isolation, samples were adjusted to have the same amount of total RNA concentration (500 pg in 7 µL). Libraries were prepared using the standard small-RNA Library Prep Kit for Illumina. To assess samples' quality before sequencing, the concentration of the samples was determined using the Fragment Analyzer and all samples met the quality requirements. Sequencing was performed using the HiSeq 4000 instrument (Illumina, San Diego, CA, USA) with 150 bp paired-ends.

4.2.3. Protein Extraction and MS/MS Sample Preparation

After RNA removal proteins were precipitated and loaded on SDS-PAGE. Image J analysis was used to enable equal loading on a subsequent SDS-PAGE (12%). Proteins were allowed to run just into the running gel before the voltage was stopped and stained with Coomassie R-250. After washing in MQ, each stained protein blob was cut from the gel as a single band and subjected to tryptic (Promega, Madison WI, USA) digestion. Peptides were extracted, desalted and dried.

Peptides were separated by an Ultimate 3000 nanoLC system (Dionex LC-Packings, Amsterdam, The Netherlands). After injection, peptides were trapped at 6 µL/min on a 10 mm × 100 µm ID trap column packed with 5 µm 120 Å ReproSil Pur C18 aqua at 2% buffer B (buffer A: 0.5% acetic acid in ultrapure water; buffer B: 80% ACN + 0.5% acetic acid in ultrapure water) and separated at 300 nL/min in a 10–40% buffer B linear gradient in 90 min (120 min inject-to-inject).

Eluting peptides were ionized into a Q Exactive mass spectrometer (Thermo Fisher, Bremen, Germany). Intact masses were measured at resolution 70.000 (at *m/z* 200) in the orbitrap using an AGC target value of 3E6 charges. The top 10 peptide signals (charge-states 2+ and higher) were submitted to MS/MS in the HCD (higher-energy collision) cell (1.6 amu isolation width, 25% normalized collision energy). MS/MS spectra were acquired at resolution 17.500 (at *m/z* 200) in the orbitrap using an AGC target value of 1E6 charges, a maxIT of 60 ms and an underfill ratio of 0.1%. Dynamic exclusion was applied with a repeat count of 1 and an exclusion time of 30 s. More details on the sample processing and for DIA method are reported in the metadata file of PXD with accession: PXD022514.

4.3. Downstream Analysis of miRNAs, mRNAs and Protein Profiles

4.3.1. Intron-RNA Processing Data

FASTQ files obtained from RNA-seq experiment were subjected to a standard pipeline as described previously [60], selecting only the spliced intron-spanning RNA reads. Gene counts were converted to counts per millions (CPM) before filtering procedures. Genes with a total count of at least 1 CPM in more than 40% of the samples were retained. Black and white cases were also retained allowing 0 values in max 18% of the positive samples group. Gene counts were then converted to TMM values through the EdgeR package in R (version 3.5.0). Differential analysis was performed with EdgeR package in R (version 3.5.0).

4.3.2. mRNA Processing Data

FASTQ files were checked for quality reads and adapters were removed with Trimmomatic [73]. Successfully quality passed reads were mapped to the human reference genome (Hg19) with STAR mapping tool [74] and gene counts were extracted with HT-Seq [75]. Normalization, quality filtering and data presence were previously described in Section 4.3.1.

4.3.3. miRNA Processing Data

cDNA libraries were sequenced at GenomeScan (Leiden, The Netherlands) and FASTQ files were acquired. Trimming, mapping and miRNA counts was performed with sRNAbench tool (version 10/14) [76]. Profiled miRNAs were then converted to CPM following the same filtering and normalization procedure for mRNA data processing. Differential analysis was performed with EdgeR package in R (version 3.5.0).

4.3.4. Protein Processing Data

MS/MS spectra were searched against the Swissprot FASTA file (release September 2015, 20197 entries, canonical and isoforms) using MaxQuant version 1.5.2.8. Peptide and protein identifications were filtered at an FDR of 1% using the decoy database strategy. The minimal peptide length was seven amino-acids. Proteins that could not be differentiated based on MS/MS spectra alone were grouped to protein groups (default MaxQuant settings). Searches were performed with the label-free quantification option selected. After contaminants removal proteins were normalized by global median. Differential analysis of spectral counts was computed in R with "ibb" package [77]. Normalized spectral counts data are reported in Table S5. Details in the data processing protocols and DIA samples are reported in the metadata file in PXD with accession PXD022514.

4.4. Differential Expression Analysis

In this study, two different statistical tests were adopted to determine the difference in the expression profiles of platelets in PDAC patients and patients with benign diseases. Differentially expressed canonical and isoforms of miRNAs were analyzed separately with the R package EdgeR (version 3.24.3) using cutoff of p-value < 0.05 and $|\text{Log2FC}| > 1$. Same parameters were applied for mRNAs and pre-mRNAs differential expression analysis. Finally, differentially expressed proteins were evaluated with the R package "ibb" and significance was determined using a threshold of p-value < 0.05.

4.5. Correlation Analysis

Correlation analyses were computed in R version 3.5.1. with cor.test() function. Significance level was set to p-value < 0.05.

4.6. Cell Culture and Transfection

Panc-1 cells were purchased from the American Type Culture Collection (ATCC, Manassas, VA, USA), while the primary cells LPC006 were isolated from laser-microdissected PDACs, as described previously [78]. Both were maintained in RPMI supplemented with 10% FCS, 1% penicillin/streptomycin, and 1% glutamine, at 37 °C under an atmosphere of 5% CO_2 in 75 cm^2 tissue culture flasks (Greiner Bio-One GmbH, Frickenhausen, Germany). When the cells were ready for transfection, they were plated the day before and then transfected with the precursor and antisense oligonucleotides (Pre-miR™ miRNA Precursor pre-miR-29a-3p and Anti-miR™ miRNA Inhibitor anti-miR-29a-3p) purchased from ThermoScientific/Ambion-Applied Biosystems, Waltham, MA, USA (Assay ID, PM12499 and AM12499, respectively) at 30 nM final concentration, as described previously [79]. Cells were plated at 5000 cells/well in 200 µL RPMI with 10% FBS and 1% antibiotics. After 24 h cells were exposed to 0.9 µL oligofectamine (Invitrogen, Paisley, UK) in serum-free medium, mixed for 10 min at room temperature, followed by the addition of 0.3 µL of 6.25 µM miR-29a-3p precursor or inhibitor. Cells were also incubated with miRNA negative controls. After overnight exposure the medium was removed from the wells and replaced with RPMI with 10% FBS, without antibiotics. Then cells were allowed to grow for additional 48 h in drug-free medium before lysis and RNA extraction to evaluate the transfection efficiency and modulation of SPARC mRNA level and cell migration.

4.7. Quantitative Real-Time PCR (RT-qPCR)

Total RNA extracted was used to perform RT-qPCR using Taqman mature miRNA primers and probes. RNA was extracted using a Trizol-chloroform protocol (Sigma, St. Louis, MO, USA). RNA yields and purity were checked by measuring optical density at 260/280nm with a Nanodrop® spectrophotometer (NanoDrop-Technologies, Wilmington, DE, USA). Briefly, both mature miRNA miR-29a-3p and SPARC mRNA expression were measured using specific primers (assay-ID 002112 and Hs00234160_m1, respectively) for complementary DNA (cDNA) synthesis followed by Taqman PCR analysis. PCR reactions were performed on a 7500HT sequence detection system (Applied Biosystems, Foster City, CA, USA), in accordance with the manufacturer's instructions. Duplicate samples and endogenous controls for miRNA and mRNA normalization (snRNA U6 and GAPDH) were used throughout. Quantification of miRNA relative expression was performed evaluating the threshold cycle (Ct) and normalized to U6, as described previously [79]. Similarly, expression levels of SPARC mRNA were normalized to GAPDH, and quantitation of gene expression was performed using the ΔΔCt calculation, where the amount of target gene, normalized to GAPDH and relative to the calibrator (untreated control cells), is given as $2^{-\Delta\Delta Ct}$. Specimens were amplified in triplicate with appropriate non-template controls, and the coefficient of variation was <1% for all replicates.

4.8. Migration Assay

For the migration assay the cells transfected with pre-miR-29a-3p or anti-miR-29a-3p or miR-negative control, as described above, were plated at a density of 2×10^4 cells/well onto 96 wells plates, and artificial wound tracks were created by scraping with a specific scratcher within the confluent monolayers. After removal of the detached cells by PBS washing, the medium was refreshed and cells ability to migrate into the wound area was assessed by comparing the pixels of the wound tracks in the images taken at the beginning of the exposure (time 0), with those taken after 8, 16, and 24 h, with the LeicaDMI300B-station integrated with Scratch Assay software (Digital-Cell Imaging Labs, Keerbergen, Belgium), as described previously [80].

4.9. Functional Pathways Enrichment

To investigate the regulatory role of miRNA in PDAC and benign disease, a GSEA was performed with fgsea R package (version 1.8.0). Differentially expressed intron-spanning reads and proteins were ranked based on their *p*-value and FC [−log10(*p*-value)*FC sign]. A compendium of Gene Ontology terms (biological processes, molecular function and cellular component) was downloaded from the Broad Institute: MSigDB [81] version 6.0, including Hallmarks of Cancer.

MiEAA web based tool [82] was used to evaluate biological processes enrichments in miRNAs. Performing an over-representation analysis, the normalized enrichment score (NES) was not given by the database and was set to 1.5 for each significant enrichment.

5. Conclusions

In the present study, we investigated the biological mechanisms acting in platelets of PDAC patients. Patients with benign disease were included in this study as a reference for inflammatory noise. Through an extensive gene ontology mining of different –omics data we demonstrated that active RNA processing, splicing signals events and translation initiation are specific terms in the biology of circulating platelets in patients with PDAC. In particular, the differential regulation of some genes, such as SPARC, in PDAC and benign platelets, makes platelets an interesting source for diagnostic tools, together with the different miRNAs and proteomic profiles. Further understanding of the real function of these profiles is essential for biomarker discovering when using a machine learning approach. The analysis of specific platelets content could therefore provide a dynamic and powerful approach for the specific diagnosis of PDAC, paving the way for early

detection intervention strategies, which represent the greatest hope for making substantial improvements in survival for this disease.

Supplementary Materials: The following are available online at https://www.mdpi.com/2072-6694/13/1/66/s1, Figure S1. Clustering of Healthy donors (HD). (a) Age distribution of benign, HD and PDAC patients included in the analysis. (b) clustering of differentially expressed intron-spanning reads between HD, PDAC and benign patients. Dashed line underlines the separation between the two main clusters. Figure S2. Age-gender match for transcriptomics and proteomics data in platelets of PDAC and benign patients. Figure S3. KM curve on plasma levels of miR-22 in PDAC patients. Figure S4. A. expected correlations with isomiRs and intron-spanning reads focused on SPTBN1. B. expected correlations with isomiRs and mRNAs focused on SPTBN1. C. differential analysis of SPTBN1 on RNA-seq data on PAAD-TCGA tissues and normal matched samples. D. Immunohistochemistry validation of SPTBN1 in resected pancreatic cancer patients. Figure S5. Modulation of miR-29a-3p levels reflecting the transfection efficiency of pre- and anti-miR-29a-3p, as assessed by qRT-PCR, 48 h post transfection. Significance was assessed with T-Student Test (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.0001$). Dashed line refers to comparative levels in miR-negative controls. Figure S6. In silico validation of protein biomarkers for PDAC diagnosis from blood platelets. Table S1. Clinicopathological characteristics of PDAC and benign patients included in this study. Table S2. KEGG analysis on SPARC-MiRNA targets for intron-spanning reads and mRNAs. Table S3. Multivariate regression of clinical features associated to SPARC and miR-29 expression. Table S4. Putative protein biomarkers resulting from proteomics analysis. Table S5. Normalized spectral counts data of matched proteomics samples.

Author Contributions: Conceptualization, G.M., L.L.M., D.K.-L., T.W. and E.G.; Data curation, G.M. and S.R.P.; Formal analysis, G.M., I.G., S.G.J.G.I.V. and R.P.; Funding acquisition, C.R.J., G.K., T.W. and E.G.; Investigation, G.M., L.L.M. and E.G.; Methodology, G.M., L.L.M., I.G., S.G.J.G.I.V., T.V.P., C.R.J., D.K.-L., T.W. and E.G.; Project administration, C.R.J., G.K., T.W. and E.G.; Resources, C.R.J., T.W. and E.G.; Software, T.V.P. and S.R.P.; Supervision, T.V.P., C.R.J., D.K.-L., T.W. and E.G.; Visualization, G.M. and D.K.-L.; Writing—original draft, G.M.; Writing—review & editing, G.M., L.L.M., I.G., S.G.J.G.I.V., M.C., T.Y.S.L.L., L.M., A.E.F., C.R.J., G.K., T.W. and E.G. All authors have read and agreed to the published version of the manuscript.

Funding: I.G: D.K-L and T.W. are members of the ELBA Consortium and received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 76592. KWF grant Dutch Cancer Society (grant#10401 (E.G., G.K. and T.W.) and #10212 (C.J., E.G.)), Italian Association for Cancer Research AIRC/Start-Up grant, Italy (E.G.), Fondazione Pisana Per La Scienza, Italy (E.G.). CCA Foundation and Bennink Foundation (E.G., T.W., G.K., L.M., T.L.L.)

Institutional Review Board Statement: The study design and protocol for the analysis of the samples were approved by the local Medical Ethical Board of the Amsterdam UMC, VU University Amsterdam (VUMC#14438, BUP-2012-05-09) in accordance with the ethical guidelines of the Declaration of Helsinki.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available in GEO repository under ID accession: GSE160252 (raw RNA-seq data); proteomeXchange under ID accession: PXD022514 (raw proteomics data); biological networks are presented here: http://platelnet.eu.ngrok.io and the R script code is available at: https://github.com/Giulia221091/Platel-net.

Acknowledgments: A special acknowledgment goes to Francois Rustenburg and Bart Kok for their meticulous work on the RNA and miRNA quality check.

Conflicts of Interest: T.W. is the inventor on relevant patent applications. T.W. received funding from Illumina, Inc and is shareholder of GRAIL, Inc. The other authors declare no conflict of interest.

Abbreviations

lv3p	length variant in 3′
lv5p	length variant in 5′
nta#A	non-templated addition of base A
nta#T	non-templated addition of base T
nta#C	non-templated addition of base C
nta#G	non-templated addition of base G
exactNucVar	exact nucleotide variant
mv	multiple variant
mlv3p	multiple length variant in 3′
mlv5p	multiple length variant in 5′

References

1. McGuigan, A.; Kelly, P.; Turkington, R.C.; Jones, C.; Coleman, H.G.; McCain, R.S. Pancreatic cancer: A review of clinical diagnosis, epidemiology, treatment and outcomes. *World J. Gastroenterol.* **2018**, *24*, 4846–4861. [CrossRef]
2. Ryan, D.P.; Hong, T.S.; Bardeesy, N. Pancreatic adenocarcinoma. *N. Engl. J. Med.* **2014**, *371*, 1039–1049. [CrossRef]
3. Giovannetti, E.; van der Borden, C.L.; Frampton, A.E.; Ali, A.; Firuzi, O.; Peters, G.J. Never let it go: Stopping key mechanisms underlying metastasis to fight pancreatic cancer. *Semin. Cancer Biol.* **2017**, *44*, 43–59. [CrossRef]
4. Rickles, F.R.; Falanga, A. Activation of clotting factors in cancer. In *Coagulation in Cancer*; Kwaan, H.C., Green, D., Eds.; Springer: Boston, MA, USA, 2009; Volume 148, pp. 31–41. ISBN 978-0-387-79961-2.
5. Ho-Tin-Noe, B.; Goerge, T.; Wagner, D.D. Platelets: Guardians of tumor vasculature. *Cancer Res.* **2009**, *69*, 5623–5626. [CrossRef]
6. Mandalà, M.; Reni, M.; Cascinu, S.; Barni, S.; Floriani, I.; Cereda, S.; Berardi, R.; Mosconi, S.; Torri, V.; Labianca, R. Venous thromboembolism predicts poor prognosis in irresectable pancreatic cancer patients. *Ann. Oncol.* **2007**, *18*, 1660–1665. [CrossRef]
7. Ishigaki, K.; Nakai, Y.; Isayama, H.; Saito, K.; Hamada, T.; Takahara, N.; Mizuno, S.; Mohri, D.; Kogure, H.; Matsubara, S.; et al. Thromboembolisms in advanced pancreatic cancer: A retrospective analysis of 475 patients. *Pancreas* **2017**, *46*, 1069–1075. [CrossRef]
8. Ansari, D.; Ansari, D.; Andersson, R.; Andrén-Sandberg, Å. Pancreatic cancer and thromboembolic disease, 150 years after trousseau. *Hepatobiliary Surg. Nutr.* **2015**, *4*, 325–335. [CrossRef]
9. Sabrkhany, S.; Kuijpers, M.J.E.; Verheul, H.M.W.; Griffioen, A.W.; Egbrink, M.G.A.O. Platelets: An unexploited data source in biomarker research. *Lancet Haematol.* **2015**, *2*, e512–e513. [CrossRef]
10. Plantureux, L.; Crescence, L.; Dignat-George, F.; Panicot-Dubois, L.; Dubois, C. Effects of platelets on cancer progression. *Thromb. Res.* **2018**, *164*, S40–S47. [CrossRef]
11. Nilsson, R.J.A.; Balaj, L.; Hulleman, E.; van Rijn, S.; Pegtel, D.M.; Walraven, M.; Widmark, A.; Gerritsen, W.R.; Verheul, H.M.; Vandertop, W.P.; et al. Blood platelets contain tumor-derived RNA biomarkers. *Blood* **2011**, *118*, 3680–3683. [CrossRef]
12. Best, M.G.; Vancura, A.; Wurdinger, T. Platelet RNA as a circulating biomarker trove for cancer diagnostics. *J. Thromb. Haemost.* **2017**, *15*, 1295–1306. [CrossRef] [PubMed]
13. Franco, A.T.; Corken, A.; Ware, J. Platelets at the interface of thrombosis, inflammation, and cancer. *Blood* **2015**, *126*, 582–588. [CrossRef] [PubMed]
14. Newman, P.J.; Gorski, J.; White, G.C.; Gidwitz, S.; Cretney, C.J.; Aster, R.H. Enzymatic amplification of platelet-specific messenger RNA using the polymerase chain reaction. *J. Clin. Investig.* **1988**, *82*, 739–743. [CrossRef] [PubMed]
15. Gnatenko, D.V.; Dunn, J.J.; Schwedes, J.; Bahou, W.F. Transcript profiling of human platelets using microarray and serial analysis of gene expression (SAGE). In *DNA and RNA Profiling in Human Blood*; Bugert, P., Ed.; Humana Press: Totowa, NJ, USA, 2009; Volume 496, pp. 245–272. ISBN 978-1-934115-93-0.
16. Nassa, G.; Giurato, G.; Cimmino, G.; Rizzo, F.; Ravo, M.; Salvati, A.; Nyman, T.A.; Zhu, Y.; Vesterlund, M.; Lehtiö, J.; et al. Splicing of platelet resident pre-MRNAs upon activation by physiological stimuli results in functionally relevant proteome modifications. *Sci. Rep.* **2018**, *8*. [CrossRef]
17. Sabrkhany, S.; Griffioen, A.W.; oude Egbrink, M.G.A. The role of blood platelets in tumor angiogenesis. *Biochim. Biophys. Acta BBA Rev. Cancer* **2011**, *1815*, 189–196. [CrossRef]
18. In 't Veld, S.G.J.G.; Wurdinger, T. Tumor-educated platelets. *Blood* **2019**, *133*, 2359–2364. [CrossRef]
19. Best, M.G.; Sol, N.; Kooi, I.; Tannous, J.; Westerman, B.A.; Rustenburg, F.; Schellen, P.; Verschueren, H.; Post, E.; Koster, J.; et al. RNA-Seq of tumor-educated platelets enables blood-based pan-cancer, multiclass, and molecular pathway cancer diagnostics. *Cancer Cell* **2015**, *28*, 666–676. [CrossRef]
20. Huang, G.; Yuan, M.; Chen, M.; Li, L.; You, W.; Li, H.; Cai, J.J.; Ji, G. Integrating multiple fitting regression and bayes decision for cancer diagnosis with transcriptomic data from tumor-educated blood platelets. *Analyst* **2017**, *142*, 3588–3597. [CrossRef]
21. Large, T.Y.S.L.; Veld, S.G.I.; Sol, N.; Best, M.G.; Mantini, G.; van der Lelij, E.; Rustenburg, F.; Verschueren, H.; Funel, N.; Garajova, I.; et al. Abstract 2596: Spliced messenger RNA of tumor-educated platelets yields a new diagnostic prospective for pancreatic cancer. In Proceedings of the Clinical Research (Excluding Clinical Trials), American Association for Cancer Research, Chicago, IL, USA, 1 July 2018; p. 2596.

22. Weyrich, A.S.; Lindemann, S.; Tolley, N.D.; Kraiss, L.W.; Dixon, D.A.; Mahoney, T.M.; Prescott, S.P.; McIntyre, T.M.; Zimmerman, G.A. Change in protein phenotype without a nucleus: Translational control in platelets. *Semin. Thromb. Hemost.* **2004**, *30*, 491–498. [CrossRef]
23. Denis, M.M.; Tolley, N.D.; Bunting, M.; Schwertz, H.; Jiang, H.; Lindemann, S.; Yost, C.C.; Rubner, F.J.; Albertine, K.H.; Swoboda, K.J.; et al. Escaping the nuclear confines: Signal-dependent Pre-MRNA splicing in anucleate platelets. *Cell* **2005**, *122*, 379–391. [CrossRef]
24. Braunschweig, U.; Barbosa-Morais, N.L.; Pan, Q.; Nachman, E.N.; Alipanahi, B.; Gonatopoulos-Pournatzis, T.; Frey, B.; Irimia, M.; Blencowe, B.J. Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res.* **2014**, *24*, 1774–1786. [CrossRef] [PubMed]
25. Sunderland, N.; Skroblin, P.; Barwari, T.; Huntley, R.P.; Lu, R.; Joshi, A.; Lovering, R.C.; Mayr, M. MicroRNA biomarkers and platelet reactivity: The clot thickens. *Circ. Res.* **2017**, *120*, 418–435. [CrossRef] [PubMed]
26. Plé, H.; Landry, P.; Benham, A.; Coarfa, C.; Gunaratne, P.H.; Provost, P. The repertoire and features of human platelet MicroRNAs. *PLoS ONE* **2012**, *7*, e50746. [CrossRef] [PubMed]
27. Edelstein, L.C.; Bray, P.F. Noncoding RNAs in platelet biology. In *Platelets in Thrombotic and Non-Thrombotic Disorders*; Gresele, P., Kleiman, N.S., Lopez, J.A., Page, C.P., Eds.; Springer International Publishing: Cham, Swizerland, 2017; pp. 239–252; ISBN 978-3-319-47460-1.
28. Koppers-Lalic, D.; Hackenberg, M.; Bijnsdorp, I.V.; van Eijndhoven, M.A.J.; Sadek, P.; Sie, D.; Zini, N.; Middeldorp, J.M.; Ylstra, B.; de Menezes, R.X.; et al. Nontemplated nucleotide additions distinguish the small RNA composition in cells from exosomes. *Cell Rep.* **2014**, *8*, 1649–1658. [CrossRef]
29. Koppers-Lalic, D.; Hackenberg, M.; de Menezes, R.; Misovic, B.; Wachalska, M.; Geldof, A.; Zini, N.; de Reijke, T.; Wurdinger, T.; Vis, A.; et al. Non-invasive prostate cancer detection by measuring MiRNA variants (IsomiRs) in urine extracellular vesicles. *Oncotarget* **2016**, *7*. [CrossRef]
30. Masud Karim, S.M.; Liu, L.; Le, T.D.; Li, J. Identification of MiRNA-MRNA regulatory modules by exploring collective group relationships. *BMC Genom.* **2016**, *17*. [CrossRef]
31. Bartel, D.P. MicroRNAs: Target recognition and regulatory functions. *Cell* **2009**, *136*, 215–233. [CrossRef]
32. Agarwal, V.; Bell, G.W.; Nam, J.-W.; Bartel, D.P. Predicting effective MicroRNA target sites in mammalian MRNAs. *eLife* **2015**, *4*. [CrossRef]
33. Eulalio, A.; Huntzinger, E.; Izaurralde, E. Getting to the root of MiRNA-mediated gene silencing. *Cell* **2008**, *132*, 9–14. [CrossRef]
34. Fabian, M.R.; Sonenberg, N. The mechanics of MiRNA-mediated gene silencing: A look under the hood of MiRISC. *Nat. Struct. Mol. Biol.* **2012**, *19*, 586–593. [CrossRef]
35. Pellegrino, L.; Stebbing, J.; Braga, V.M.; Frampton, A.E.; Jacob, J.; Buluwela, L.; Jiao, L.R.; Periyasamy, M.; Madsen, C.D.; Caley, M.P.; et al. MiR-23b Regulates cytoskeletal remodeling, motility and metastasis by directly targeting multiple transcripts. *Nucleic Acids Res.* **2013**, *41*, 5400–5412. [CrossRef] [PubMed]
36. Seo, J.; Jin, D.; Choi, C.-H.; Lee, H. Integration of MicroRNA, MRNA, and protein expression data for the identification of cancer-related MicroRNAs. *PLOS ONE* **2017**, *12*, e0168412. [CrossRef] [PubMed]
37. Best, M.G.; Veld, S.G.J.G.I.; Sol, N.; Wurdinger, T. RNA sequencing and swarm intelligence–Enhanced classification algorithm development for blood-based disease diagnostics using spliced blood platelet RNA. *Nat. Protoc.* **2019**, *14*, 1206–1234. [CrossRef] [PubMed]
38. Tian, C.; Clauser, K.R.; Öhlund, D.; Rickelt, S.; Huang, Y.; Gupta, M.; Mani, D.R.; Carr, S.A.; Tuveson, D.A.; Hynes, R.O. Proteomic analyses of ECM during pancreatic ductal adenocarcinoma progression reveal different contributions by tumor and stromal cells. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 19609–19618. [CrossRef]
39. Iozzo, R.V.; Sanderson, R.D. Proteoglycans in cancer biology, tumour microenvironment and angiogenesis. *J. Cell. Mol. Med.* **2011**, *15*, 1013–1031. [CrossRef]
40. Jiang, N.; Dai, Q.; Su, X.; Fu, J.; Feng, X.; Peng, J. Role of PI3K/AKT Pathway in cancer: The framework of malignant behavior. *Mol. Biol. Rep.* **2020**, *47*, 4587–4629. [CrossRef]
41. Vara, J.Á.F.; Casado, E.; de Castro, J.; Cejas, P.; Belda-Iniesta, C.; González-Barón, M. PI3K/Akt Signalling pathway and cancer. *Cancer Treat. Rev.* **2004**, *30*, 193–204. [CrossRef]
42. Ahmed, S.; Bradshaw, A.-D.; Gera, S.; Dewan, M.; Xu, R. The TGF-β/Smad4 signaling pathway in pancreatic carcinogenesis and Its clinical significance. *J. Clin. Med.* **2017**, *6*, 5. [CrossRef]
43. Alvarez, M.A.; Freitas, J.P.; Mazher Hussain, S.; Glazer, E.S. TGF-β inhibitors in metastatic pancreatic ductal adenocarcinoma. *J. Gastrointest. Cancer* **2019**, *50*, 207–213. [CrossRef]
44. Cimmino, G.; Tarallo, R.; Nassa, G.; Filippo, M.R.D.; Giurato, G.; Ravo, M.; Rizzo, F.; Conte, S.; Pellegrino, G.; Cirillo, P.; et al. Activating stimuli induce platelet MicroRNA Modulation and proteome reorganisation. *Thromb. Haemost.* **2015**, *114*, 96–108. [CrossRef]
45. Meijer, L.L.; Puik, J.R.; Le Large, T.Y.S.; Heger, M.; Dijk, F.; Funel, N.; Wurdinger, T.; Garajová, I.; van Grieken, N.C.T.; van de Wiel, M.A.; et al. Unravelling the diagnostic dilemma: A MicroRNA panel of circulating MiR-16 and MiR-877 as a diagnostic classifier for distal bile duct tumors. *Cancers* **2019**, *11*, 1181. [CrossRef] [PubMed]
46. Zhu, J.; Nie, S.; Wu, J.; Lubman, D.M. Target proteomic profiling of frozen pancreatic CD24+ adenocarcinoma tissues by immuno-laser capture microdissection and Nano-LC–MS/MS. *J. Proteome Res.* **2013**, *12*, 2791–2804. [CrossRef] [PubMed]

47. Reduced expression of the membrane skeleton protein beta1-spectrin (SPTBN1) Is associated with worsened prognosis in pancreatic cancer. *Histol. Histopathol.* **2010**, 1497–1506. [CrossRef]
48. Mantini, G.; Vallés, A.M.; Le Large, T.Y.S.; Capula, M.; Funel, N.; Pham, T.V.; Piersma, S.R.; Kazemier, G.; Bijlsma, M.F.; Giovannetti, E.; et al. Co-expression analysis of pancreatic cancer proteome reveals biology and prognostic biomarkers. *Cell. Oncol.* **2020**. [CrossRef] [PubMed]
49. Sengupta, S.; den Boon, J.A.; Chen, I.-H.; Newton, M.A.; Stanhope, S.A.; Cheng, Y.-J.; Chen, C.-J.; Hildesheim, A.; Sugden, B.; Ahlquist, P. MicroRNA 29c is down-regulated in nasopharyngeal carcinomas, up-regulating MRNAs encoding extracellular matrix proteins. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 5874–5878. [CrossRef] [PubMed]
50. Russ, A.C.; Sander, S.; Luck, S.C.; Lang, K.M.; Bauer, M.; Rucker, F.G.; Kestler, H.A.; Schlenk, R.F.; Dohner, H.; Holzmann, K.; et al. Integrative nucleophosmin mutation-associated MicroRNA and gene expression pattern analysis identifies novel micrornA—Target gene interactions in acute myeloid leukemia. *Haematologica* **2011**, *96*, 1783–1791. [CrossRef] [PubMed]
51. Wang, C.; Gao, C.; Zhuang, J.-L.; Ding, C.; Wang, Y. A combined approach identifies three MRNAs that are down-regulated by MicroRNA-29b and promote invasion ability in the breast cancer cell line MCF-7. *J. Cancer Res. Clin. Oncol.* **2012**, *138*, 2127–2136. [CrossRef]
52. Plaisier, C.L.; Pan, M.; Baliga, N.S. A MiRNA-regulatory network explains how dysregulated MiRNAs perturb oncogenic processes across diverse cancers. *Genome Res.* **2012**, *22*, 2302–2314. [CrossRef]
53. Zhu, X.-C.; Dong, Q.-Z.; Zhang, X.-F.; Deng, B.; Jia, H.-L.; Ye, Q.-H.; Qin, L.-X.; Wu, X.-Z. MicroRNA-29a Suppresses cell proliferation by targeting SPARC in hepatocellular carcinoma. *Int. J. Mol. Med.* **2012**, *30*, 1321–1326. [CrossRef]
54. Pham, T.V.; Henneman, A.A.; Jimenez, C.R. Iq: An R Package to estimate relative protein abundances from ion quantification in DIA-MS-Based Proteomics. *Bioinformatics* **2020**, *36*, 2611–2613. [CrossRef]
55. Brown, P.O.; Botstein, D. Exploring the new world of the genome with DNA microarrays. *Nat. Genet.* **1999**, *21*, 33–37. [CrossRef] [PubMed]
56. Koukourakis, M.I.; Giatromanolaki, A.; Brekken, R.A.; Sivridis, E.; Gatter, K.C.; Harris, A.L.; Sage, E.H. Enhanced expression of SPARC/Osteonectin in the tumor-associated stroma of non-small cell lung cancer is correlated with markers of hypoxia/acidity and with poor prognosis of patients. *Cancer Res.* **2003**, *63*, 5376. [PubMed]
57. Sato, N.; Fukushima, N.; Maehara, N.; Matsubayashi, H.; Koopmann, J.; Su, G.H.; Hruban, R.H.; Goggins, M. SPARC/Osteonectin is a frequent target for aberrant methylation in pancreatic adenocarcinoma and a mediator of tumor–stromal interactions. *Oncogene* **2003**, *22*, 5021–5030. [CrossRef]
58. Yunker, C.K.; Golembieski, W.; Lemke, N.; Schultz, C.R.; Cazacu, S.; Brodie, C.; Rempel, S.A. SPARC-induced increase in glioma matrix and decrease in vascularity are associated with reduced VEGF expression and secretion. *Int. J. Cancer* **2008**, *122*, 2735–2743. [CrossRef] [PubMed]
59. Motamed, K. SPARC (Osteonectin/BM-40). *Int. J. Biochem. Cell Biol.* **1999**, *31*, 1363–1366. [CrossRef]
60. Best, M.G.; Sol, N.; In 't Veld, S.G.J.G.; Vancura, A.; Muller, M.; Niemeijer, A.-L.N.; Fejes, A.V.; Tjon Kon Fat, L.-A.; Huis In 't Veld, A.E.; Leurs, C.; et al. Swarm intelligence-enhanced detection of non-small-cell lung cancer using tumor-educated platelets. *Cancer Cell* **2017**, *32*, 238–252.e9. [CrossRef] [PubMed]
61. Gagliardi, D.; Dziembowski, A. 5′ and 3′ modifications controlling RNA degradation: From safeguards to executioners. *Philos. Trans. R. Soc. B Biol. Sci.* **2018**, *373*, 20180160. [CrossRef]
62. Houseley, J.; Tollervey, D. The Many pathways of RNA degradation. *Cell* **2009**, *136*, 763–776. [CrossRef]
63. Cohen, J.D.; Javed, A.A.; Thoburn, C.; Wong, F.; Tie, J.; Gibbs, P.; Schmidt, C.M.; Yip-Schneider, M.T.; Allen, P.J.; Schattner, M.; et al. Combined circulating tumor DNA and protein biomarker-based liquid biopsy for the earlier detection of pancreatic cancers. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 10202–10207. [CrossRef]
64. Gall, T.M.H.; Belete, S.; Khanderia, E.; Frampton, A.E.; Jiao, L.R. Circulating tumor cells and cell-free DNA in pancreatic ductal adenocarcinoma. *Am. J. Pathol.* **2019**, *189*, 71–81. [CrossRef]
65. Liu, X.-D.; Wu, H.; Li, Y.; Liu, X.; Zhang, Z.; Yu, L.; Qin, Z.; Su, Z.; Liu, R.; He, Q.; et al. Early detection of pancreatic ductal adenocarcinoma using methylation signatures in circulating tumour DNA. *Ann. Oncol.* **2019**, *30*, v261–v262. [CrossRef]
66. Lehmann-Werman, R.; Neiman, D.; Zemmour, H.; Moss, J.; Magenheim, J.; Vaknin-Dembinsky, A.; Rubertsson, S.; Nellgård, B.; Blennow, K.; Zetterberg, H.; et al. Identification of tissue-specific cell death using methylation patterns of circulating DNA. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E1826–E1834. [CrossRef] [PubMed]
67. Sun, K.; Jiang, P.; Chan, K.C.A.; Wong, J.; Cheng, Y.K.Y.; Liang, R.H.S.; Chan, W.; Ma, E.S.K.; Chan, S.L.; Cheng, S.H.; et al. Plasma DNA tissue mapping by genome-wide methylation sequencing for noninvasive prenatal, cancer, and transplantation assessments. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E5503–E5512. [CrossRef] [PubMed]
68. Singh, N.; Gupta, S.; Rashid, S.; Rashid, S.; Dash, N.R.; Saraya, A. Tu2033—Quantitation of methylation load of tumor suppressor gene promoter methylation in pancreatic cancer. *Gastroenterology* **2019**, *156*, S-1176. [CrossRef]
69. Park, J.W.; Baek, I.H.; Kim, Y.T. Preliminary study analyzing the methylated genes in the plasma of patients with pancreatic cancer. *Scand. J. Surg.* **2012**, *101*, 38–44. [CrossRef]
70. Sol, N.; in 't Veld, S.G.J.G.; Vancura, A.; Tjerkstra, M.; Leurs, C.; Rustenburg, F.; Schellen, P.; Verschueren, H.; Post, E.; Zwaan, K.; et al. Tumor-educated platelet RNA for the detection and (Pseudo)progression monitoring of glioblastoma. *Cell Rep. Med.* **2020**, *1*, 100101. [CrossRef]

71. Sol, N.; Leurs, C.E.; Veld, S.G.I.T.; Strijbis, E.M.; Vancura, A.; Schweiger, M.W.; Teunissen, C.E.; Mateen, F.J.; Tannous, B.A.; Best, M.G.; et al. Blood platelet RNA enables the detection of multiple sclerosis. *Mult. Scler. J. Exp. Transl. Clin.* **2020**, *6*. [CrossRef]
72. Heinhuis, K.M.; In 't Veld, S.G.J.G.; Dwarshuis, G.; van den Broek, D.; Sol, N.; Best, M.G.; van Coevorden, F.; Haas, R.L.; Beijnen, J.H.; van Houdt, W.J.; et al. RNA-sequencing of tumor-educated platelets, a novel biomarker for blood-based sarcoma diagnostics. *Cancers* **2020**, *12*, 1372. [CrossRef]
73. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef]
74. Dobin, A.; Davis, C.A.; Zaleski, C.; Schlesinger, F.; Drenkow, J.; Chaisson, M.; Batut, P.; Jha, S.; Gingeras, T.R. STAR: Ultrafast universal RNA-Seq aligner. *Bioinformatics* **2012**, *29*, 15–21. [CrossRef]
75. Anders, S.; Pyl, P.T.; Huber, W. HTSeq—A python framework to work with high-throughput sequencing data. *Bioinformatics* **2015**, *31*, 166–169. [CrossRef] [PubMed]
76. Rueda, A.; Barturen, G.; Lebrón, R.; Gómez-Martín, C.; Alganza, Á.; Oliver, J.L.; Hackenberg, M. SRNAtoolbox: An integrated collection of small RNA research tools. *Nucleic Acids Res.* **2015**, *43*, W467–W473. [CrossRef] [PubMed]
77. Pham, T.V.; Piersma, S.R.; Warmoes, M.; Jimenez, C.R. On the beta-binomial model for analysis of spectral count data in label-free tandem mass spectrometry-based proteomics. *Bioinformatics* **2010**, *26*, 363–369. [CrossRef] [PubMed]
78. Giovannetti, E.; Funel, N.; Peters, G.J.; Del Chiaro, M.; Erozenci, L.A.; Vasile, E.; Leon, L.G.; Pollina, L.E.; Groen, A.; Falcone, A.; et al. MicroRNA-21 in pancreatic cancer: Correlation with clinical outcome and pharmacologic aspects underlying its role in the modulation of gemcitabine activity. *Cancer Res.* **2010**, *70*, 4528–4538. [CrossRef]
79. Giovannetti, E.; van der Velde, A.; Funel, N.; Vasile, E.; Perrone, V.; Leon, L.G.; de Lio, N.; Avan, A.; Caponi, S.; Pollina, L.E.; et al. High-Throughput MicroRNA (MiRNAs) arrays unravel the prognostic role of MiR-211 in pancreatic cancer. *PLoS ONE* **2012**, *7*, e49145. [CrossRef]
80. Giovannetti, E.; Wang, Q.; Avan, A.; Funel, N.; Lagerweij, T.; Lee, J.-H.; Caretti, V.; van der Velde, A.; Boggi, U.; Wang, Y.; et al. Role of CYB5A in pancreatic cancer prognosis and autophagy modulation. *JNCI J. Natl. Cancer Inst.* **2014**, *106*. [CrossRef]
81. Subramanian, A.; Tamayo, P.; Mootha, V.K.; Mukherjee, S.; Ebert, B.L.; Gillette, M.A.; Paulovich, A.; Pomeroy, S.L.; Golub, T.R.; Lander, E.S.; et al. Gene Set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15545. [CrossRef]
82. Backes, C.; Khaleeq, Q.T.; Meese, E.; Keller, A. MiEAA: MicroRNA enrichment analysis and annotation. *Nucleic Acids Res.* **2016**, *44*, W110–W116. [CrossRef]

Supplementary Appendix Chapter 7

Omics Analysis of Educated Platelets in Cancer and Benign Disease of the Pancreas

Table S1. Clinicopathological characteristics of the included patients with PDAC and Benign disease.

	PDAC	**Benign**	p-value*
	(N=11)	(N=11)	
Age-yrs			0.195
Mean (SD)	69 (±10)	62 (±14)	
Sex- No. (%)			0.521
Male	5 (45)	2 (18)	
Female	6 (55)	9 (82)	
Stage° -No. (%)			–
IA	1 (9)	–	
IIA	4 (36)	–	
IIB	6 (55)	–	
Benign diagnosis			–
Cholelithiasis	–	6 (54)	
Fibrosis of the pancreas	–	2 (18)	
Pancreatic pseudocyst	–	1 (0.09)	
Gall bladder adenoma	–	2 (18)	
CA19-9 - No. (%)			**0.006****
Normal (0-37 U/mL)	3 (27)	9 (81)	
ULN to ≤ 100	3 (27)	1 (0.09)	
ULN to > 100	5 (45)	–	
Missing	–	1 (0.09)	
CA19-9 (U/mL)			**0.038***
Median (SD)	100 (±3254)	7 (±24)	
Bilirubin - No. (%)			**0.035***
High	5 (45)	–	
Low	6 (55)	11 (100)	
Bilirubin (µmol/L)			**0.009***
Median (SD)	14 (±49)	6 (±2)	

Table S2. KEGG analysis on SPARC-MiRNA targets for intron-spanning reads and mRNAs.

KEGG pathway	p-value	#genes	#miRNAs
ECM-receptor interaction	1.06E-86	25	4
Amoebiasis	2.47E-10	24	4
Proteoglycans in cancer	1.92E-07	32	4
Protein digestion and absorption	5.23E-07	28	4
Lysine degradation	1.51E-06	10	4
PI3K-Akt signaling pathway	2.66E-05	60	4
Platelet activation	0.00012137	28	4
Focal adhesion	0.00028824	40	4
Small cell lung cancer	0.00365586	20	4
mTOR signaling pathway	0.01116579	15	4
Thyroid hormone signaling pathway	0.01370714	19	4
Prolactin signaling pathway	0.04336159	12	4
Glioma	1.03E-06	17	3
Melanoma	0.00560823	16	3
Pancreatic cancer	0.03165216	11	3
Non-small cell lung cancer	0.03370904	11	3

Table S3. Multivariate regression of clinical features associated to SPARC and miR-29 expression.

	N	*p*-value (SPARC)	*p*-value (miR-29)
Age		0.88	0.28
Low (<65)	9		
High (>=65)	13		
Sex		0.06	0.11
Male	7		
Female	15		
Stage		0.59	0.11
I-IIA	5		
IIB	6		

Table S4. Putative protein biomarkers resulting from proteomics analysis.

	Proteomics	**Transcriptomics**		**Small-RNA**	
Protein Biomarker candidate	Proteomics	Intron-spanning RNA	mRNA	miRNA-canonical	IsomiRs
AGT	UP Benign	not detected	not detected	DE: miR-26b-5p (UP PDAC)	PDAC: mlv5p,lv3p,lv5p,nta#A
HBA1	UP PDAC	no difference	no difference	no miRNA targets	
HBD	UP PDAC	no difference	no difference	DE: miR-26b-5p (UP PDAC)*	PDAC: mlv5p,lv3p,lv5p,nta#A*
PRDX2	UP PDAC	no difference	no difference	no difference	
CA1	UP PDAC	no difference	no difference	DE: miR-26b-5p; miR-132-3p (UP PDAC)*	PDAC: mlv5p,lv3p,lv5p,nta#A*

Table S5. Normalized spectral counts data of matched proteomics samples used in this study.

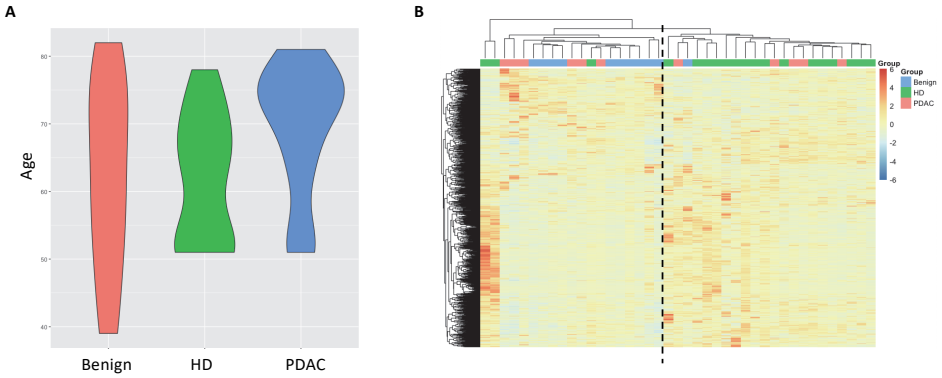


Figure S1. Clustering of Healthy donors (HD). (a) Age distribution of benign, HD and PDAC patients included in the analysis. (b) clustering of differentially expressed intron-spanning reads between HD, PDAC and benign patients. Dashed line underlines the separation between the two main clusters.

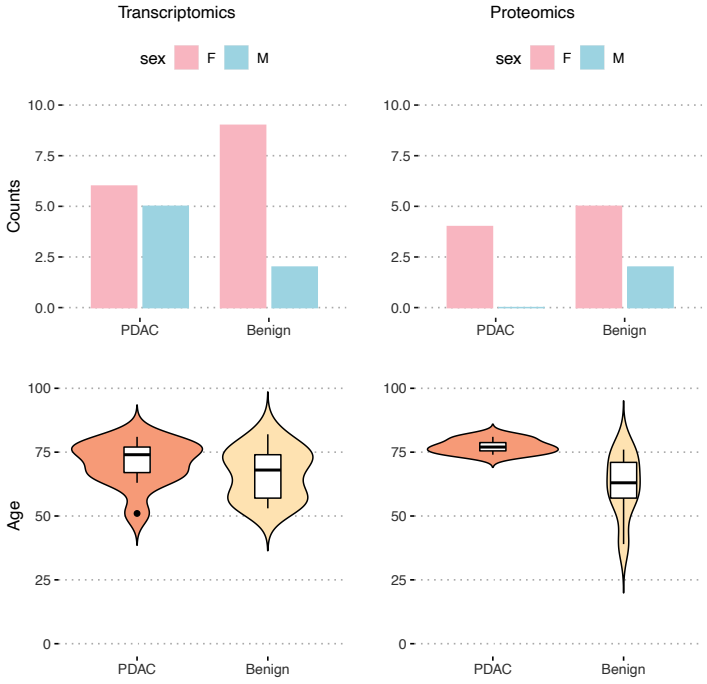


Figure S2. Age-gender match for transcriptomics and proteomics data in platelets of PDAC and benign patients.

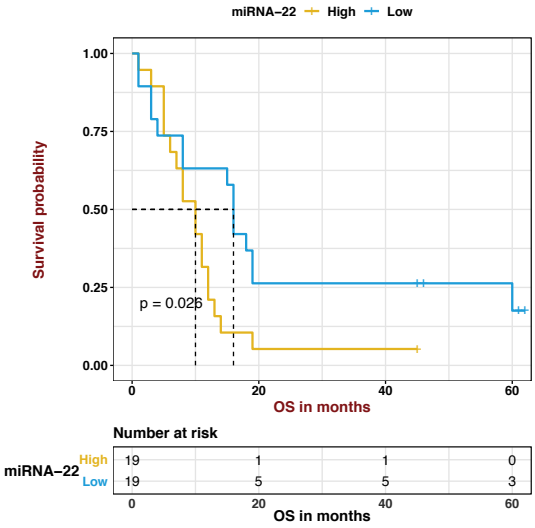


Figure S3. KM curve on plasma levels of miR-22 in PDAC patients.

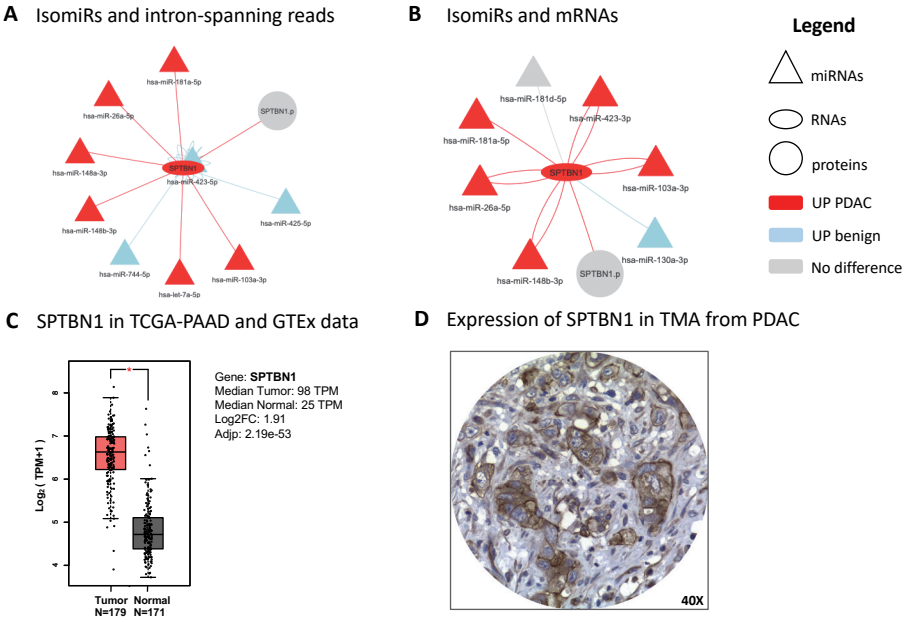


Figure S4. A. expected correlations with isomiRs and intron-spanning reads focused on SPTBN1. B. expected correlations with isomiRs and mRNAs focused on SPTBN1. C. differential analysis of SPTBN1 on RNA-seq data on PAAD-TCGA tissues and normal matched samples. D. Immunohistochemistry validation of SPTBN1 in resected pancreatic cancer patients.

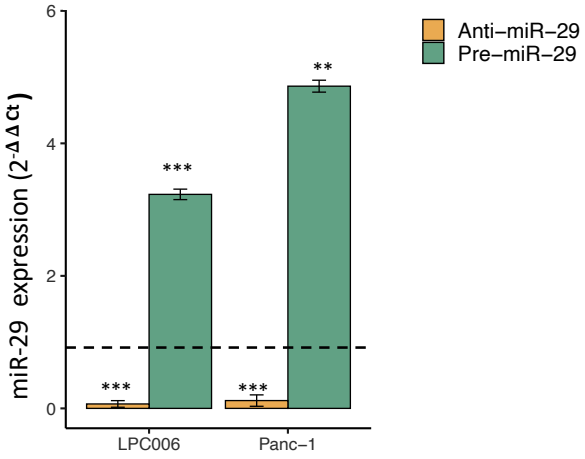
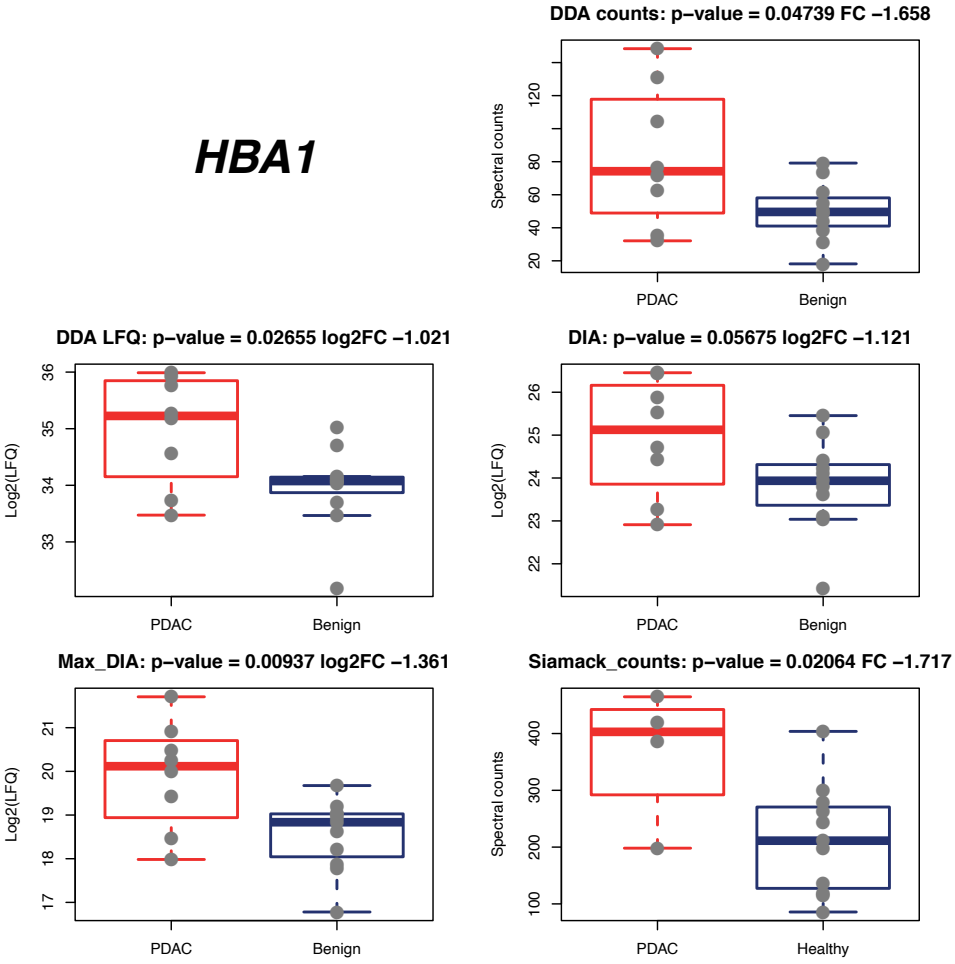
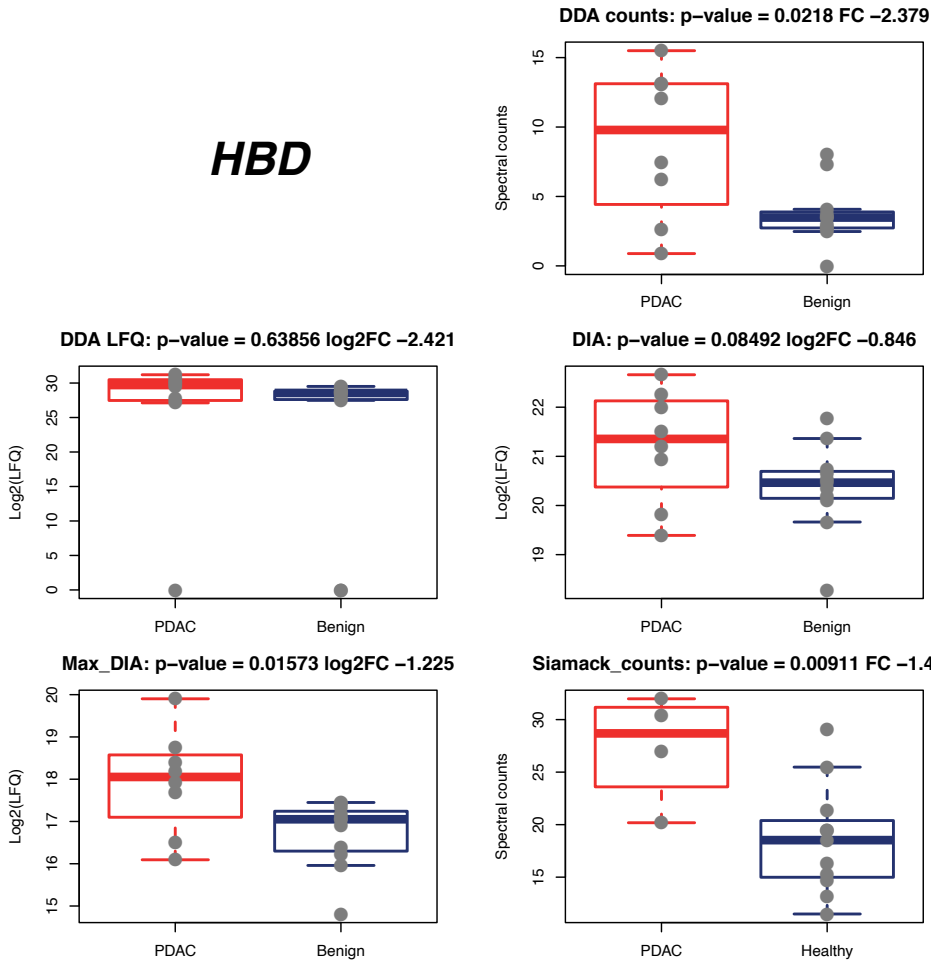
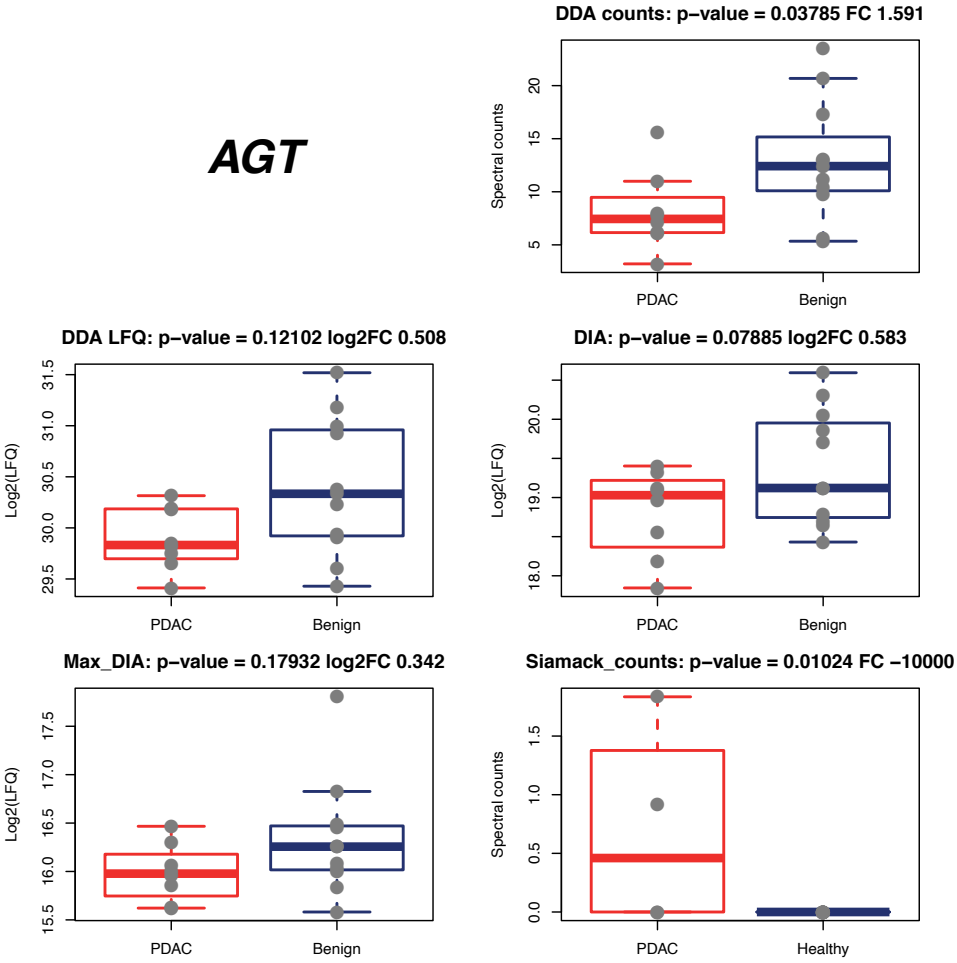


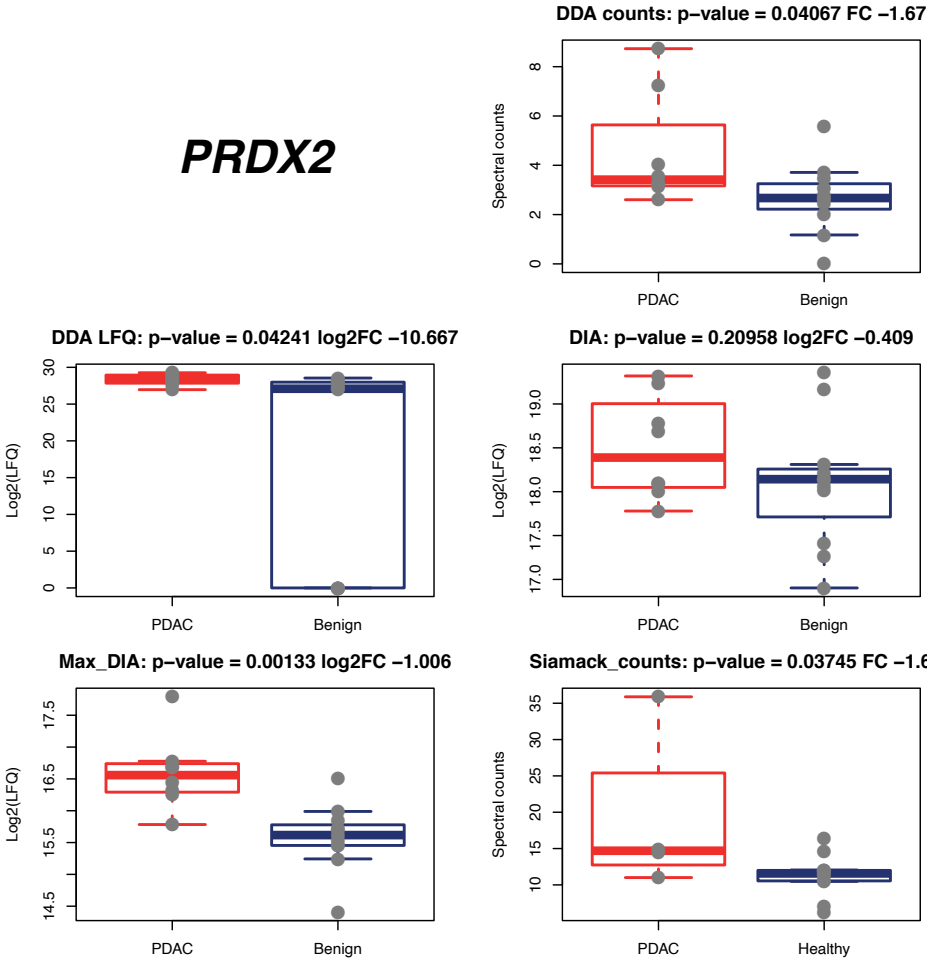
Figure S5. Modulation of miR-29a-3p levels reflecting the transfection efficiency of pre- and anti-miR-29a-3p, as assessed by qRT-PCR, 48 h post transfection. Significance was assessed with T-Student Test (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.0001$). Dashed line refers to comparative levels in miR-negative controls.


HBA1
DDA counts: p–value = 0.04739 FC –1.658
Spectral counts
20
40
60
80
120
PDAC
Benign
DDA LFQ: p–value = 0.02655 log2FC –1.021
Log2(LFQ)
33
34
35
36
PDAC
Benign
DIA: p–value = 0.05675 log2FC –1.121
Log2(LFQ)
22
23
24
25
26
PDAC
Benign
Max_DIA: p–value = 0.00937 log2FC –1.361
Log2(LFQ)
17
18
19
20
21
PDAC
Benign
Siamack_counts: p–value = 0.02064 FC –1.717
Spectral counts
100
200
300
400
PDAC
Healthy


HBD
DDA counts: p–value = 0.0218 FC –2.379
Spectral counts
PDAC
Benign
DDA LFQ: p–value = 0.63856 log2FC –2.421
Log2(LFQ)
PDAC
Benign
DIA: p–value = 0.08492 log2FC –0.846
Log2(LFQ)
PDAC
Benign
Max_DIA: p–value = 0.01573 log2FC –1.225
Log2(LFQ)
PDAC
Benign
Siamack_counts: p–value = 0.00911 FC –1.475
Spectral counts
PDAC
Healthy


AGT
DDA counts: p–value = 0.03785 FC 1.591
Spectral counts
5
10
15
20
PDAC
Benign
DDA LFQ: p–value = 0.12102 log2FC 0.508
Log2(LFQ)
29.5
30.0
30.5
31.0
31.5
PDAC
Benign
DIA: p–value = 0.07885 log2FC 0.583
Log2(LFQ)
18.0
19.0
20.0
PDAC
Benign
Max_DIA: p–value = 0.17932 log2FC 0.342
Log2(LFQ)
15.5
16.0
16.5
17.0
17.5
PDAC
Benign
Siamack_counts: p–value = 0.01024 FC –10000
Spectral counts
0.0
0.5
1.0
1.5
PDAC
Healthy

PRDX2


DDA counts: p–value = 0.04067 FC –1.67
Spectral counts
PDAC
Benign
DDA LFQ: p–value = 0.04241 log2FC –10.667
Log2(LFQ)
PDAC
Benign
DIA: p–value = 0.20958 log2FC –0.409
Log2(LFQ)
PDAC
Benign
Max_DIA: p–value = 0.00133 log2FC –1.006
Log2(LFQ)
PDAC
Benign
Siamack_counts: p–value = 0.03745 FC –1.689
Spectral counts
PDAC
Healthy

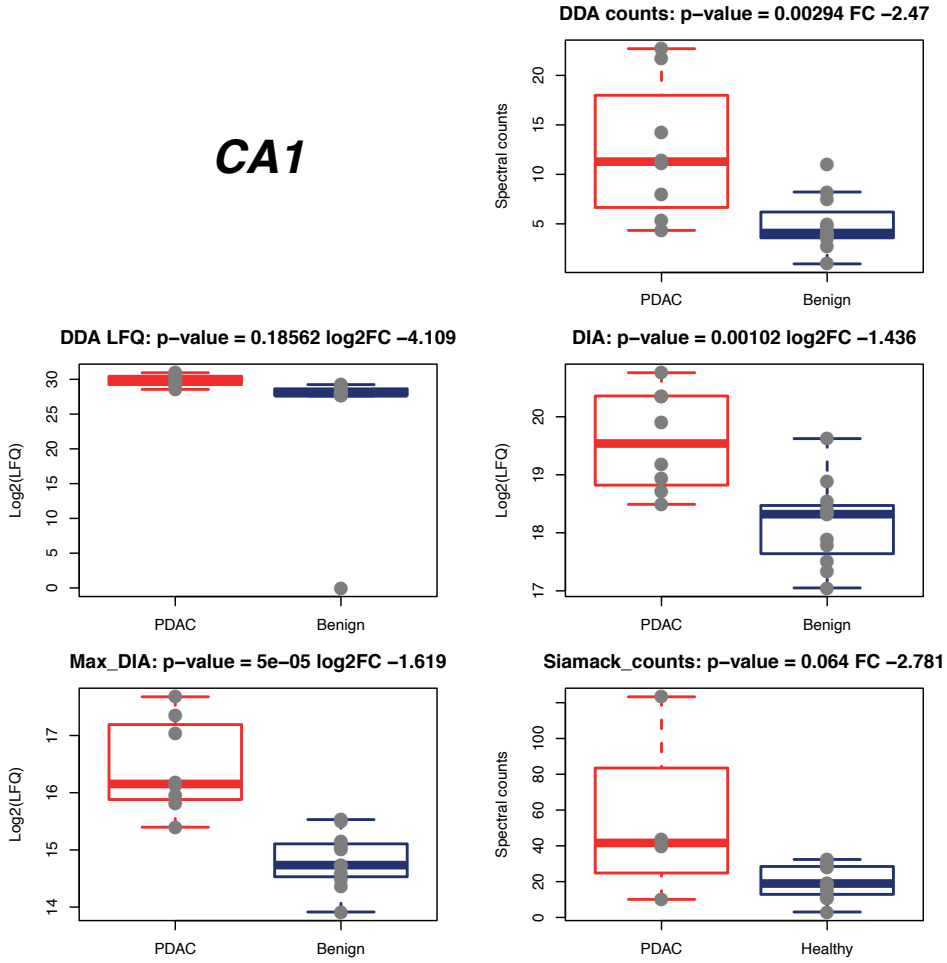


Figure S6. In silico validation of protein biomarkers for PDAC diagnosis from blood platelets.

PART THREE

PHARMACOLOGICAL STUDIES

Chapter 8

To Combine or Not Combine: Drug Interactions and Tools for Their Analysis. Reflections from the EORTC-PAMM Course on Preclinical and Early-phase Clinical Pharmacology

El Hassouni B, Mantini G, Li Petri G, Capula M, Boyd L, Weinstein HNW, Vallés-Marti A, Kouwenhoven MCM, Giovannetti E, Westerman BA, Peters GJ; EORTC PAMM Group

Anticancer Research, 2019

ANTICANCER RESEARCH *39*: 3303-3309 (2019)
doi:10.21873/anticanres.13472

Review

To Combine or Not Combine: Drug Interactions and Tools for Their Analysis. Reflections from the EORTC-PAMM Course on Preclinical and Early-phase Clinical Pharmacology

BTISSAME EL HASSOUNI[1], GIULIA MANTINI[1], GIOVANNA LI PETRI[1], MJRIAM CAPULA[2], LENKA BOYD[1], HANNAH N.W. WEINSTEIN[1], ANDREA VALLÉS-MARTI[1], MATHILDE C.M. KOUWENHOVEN[3], ELISA GIOVANNETTI[1], BART A. WESTERMAN[4] and GODEFRIDUS J. PETERS[1], on Behalf of the EORTC PAMM Group

[1]*Department of Medical Oncology, Amsterdam UMC, VU University Medical Center, Amsterdam, the Netherlands;*
[2]*Fondazione Pisana per la Scienza, Pisa, Italy;*
[3]*Department of Neurology, Brain Tumor Center Amsterdam, Amsterdam University Medical Center, Cancer Center Amsterdam, Amsterdam, the Netherlands;*
[4]*Department of Neurosurgery, Amsterdam UMC, VU University Medical Center, Amsterdam, the Netherlands*

Abstract. *Combination therapies are used in the clinic to achieve cure, better efficacy and to circumvent resistant disease in patients. Initial assessment of the effect of such combinations, usually of two agents, is frequently performed using in vitro assays. In this review, we give a short summary of the types of analyses that were presented during the Preclinical and Early-phase Clinical Pharmacology Course of the Pharmacology and Molecular Mechanisms Group, European Organization for Research and Treatment on Cancer, that can be used to determine the efficacy of drug combinations. The effect of a combination treatment can be calculated using mathematical equations based on either the Loewe additivity or Bliss independence model, or a combination of both, such as Chou and Talalay's median-drug effect model. Interactions can be additive, synergistic (more than additive), or antagonistic (less than additive). Software packages CalcuSyn (also available as CompuSyn) and Combenefit are designed to calculate the extent of the combined effects. Interestingly, the application of machine-learning methods in the prediction of combination treatments, which can include pharmacogenomic, genetic, metabolomic and proteomic profiles, might contribute to further refinement of combination regimens. However, more research is needed to apply appropriate rules of machine learning methods to ensure correct predictive models.*

Correspondence to: Professor. Dr. G.J. Peters, Department of Medical Oncology, Amsterdam UMC, VU University Medical Center, De Boelelaan 1117, 1081 HV Amsterdam, the Netherlands. Tel: +31 204442633, Fax: +31 204443844, e-mail: gj.peters@vumc.nl

Key Words: Combination treatment, synergy, Calcusyn, Compusyn, review.

Even as early as the 1960s, the majority of clinical treatments consisted of combination regimens. Combinations such as mechlorethamine, vincristine, procarbazine and prednisone (MOPP), and cyclophosphamide, hydroxydaunorubicin and oncovin with prednisone (CHOP) represented a breakthrough in the cure of lymphoma, while other combinations led to a high curation rate in childhood leukaemia (1). Depending on the type of combination used, the treatment rationale is to i) increase the efficacy of each separate drug without increasing toxicity, ii) add a drug which offers protection against toxicity, iii) bypass resistance development, or iv) target different subpopulations in a heterogeneous tumour. The initial clinical rationale was to achieve a better therapeutic effect (*e.g.* a complete response) than accomplished by each drug separately (*e.g.* only a partial response) (2). Historically, the selection of drugs to apply in combination therapies was based on the observation that each of the drugs showed antitumor activity against a certain tumour type, preferably with different toxicities of the two drugs. Doses and schedules were determined by trial and error. Soon thereafter, a complementary scientific approach was used to select combinations based on the mechanisms of action of each drug (3). An excellent example is the gemcitabine–cisplatin combination, which was initially developed by our group (4) (with the aim of preventing repair of DNA–platinum adducts) and is now standard therapy for tumours such as non-small cell lung cancer and bladder cancer. Another combination is

ANTICANCER RESEARCH *39*: 3303-3309 (2019)

5-fluorouracil (5-FU) and leucovorin (folinic acid) for which we demonstrated in model systems (cell lines, experimental tumours and in tumours of patients) that leucovorin increased and prolonged the inhibition of the 5-FU target, thymidylate synthase (5). 5-FU and leucovorin are part of the standard combination of drugs used in regimens of folinic acid, 5-FU and oxaliplatin (FOLFOX), folinic acid, 5-FU and irinotecan (FOLFIRI) (colon cancer) and FOLFIRI with oxaliplatin (FOLFIRINOX) (pancreatic cancer). The increased toxicity which is often observed with these combinations is usually controlled by combinations with anti-emetics or a corticosteroid such as dexamethasone. The latter may also have (or influence) antitumor effects (6).

Currently most combinations are established using various *in vitro* assays either focusing on the interaction of drugs on a specific target in a cell-free system, or using a pharmacological assay as summarized previously (7-9). However, an often observed mistake is the lack of proper controls (simply testing the effect of the single agent and combined effect over the whole tested concentration range on cells).

The effect of combination treatment compared to monotherapies can be described as synergistic, additive or antagonistic. The definition is dependent on the mathematical model used, but in general it can be stated that additivity means that the predicted effect of the combination is equivalent to either the sum or the product of each separate effect. Synergism is better than the expected theoretical effect (higher than the sum or lower than the product) and antagonism is worse than the expected theoretical effect (lower than the sum or higher than the product).

In this review, we give a short summary of the types of analyses that can be used to determine the efficacy of drug combinations. Furthermore, we summarize the advantages and disadvantages of these methods and lastly discuss emerging computational approaches.

Methods for Determining the Effect of Drug Combinations

Several mathematical models were initially used to evaluate drug interactions in cell-free systems, in which the definition of the reference state (meaning no interaction) was the basis. In the Loewe additivity model (10), it was hypothesized that when drug A is combined with itself, the effect would be the sum of: A+A=2A. When another drug was used, the reference state would be A+B=2A. In the Bliss independence model (11), which is most commonly used, the reference state (additivity) is a product of the fractional response, in which 0.5×0.5=0.25. Almost all current models are either a modified use of the Bliss and Loewe models, or are predominantly based on the Bliss model. Application of cell-free models to cellular systems assumes a sigmoidal dose–response curve based on the Hill equation allowing (a) fractional-effect analysis, (b) isobolograms, (c) the response surface area model, based on a mixed Loewe-Bliss; and (d) median effect analysis (Figure 1).

Fractional-effect Analysis

Fractional-effect analysis determines the theoretical additive effect of a combination by multiplying the effect of each drug alone (12). When drugs A and B are combined at an equitoxic concentration, *e.g.* achieving 50% growth inhibition (IC_{50}), the theoretical fractional effect is 0.5 for each drug, and additivity is 0.5×0.5=0.25. When drug C has a moderate effect, *e.g.* 25% growth inhibition at a specific concentration (IC_{25}), the fractional effect (*fa*) is calculated as: (*fa*)=(1-g Growth inhibition in %)/100 resulting in a value of 0.75. When treatment with drug D has only a minor effect *e.g.* 2% growth inhibition, the fractional effect is 0.98. The additivity of the combination of drugs C (fa=0.75) and A (fa=0.5) is equal to their product, *i.e.* 0.375, and that for drugs D (fa=0.98) and A (fa=0.5) is similarly computed to give 0.49. Synergism is achieved for these combinations when the experimentally determined fractional effect is lower than 0.25 (A combined with B), 0.375 (A combined with C) or (A combined with D) 0.49, respectively. Antagonism is achieved when these values are higher than 0.25, 0.375 or 0.49, respectively. This method is rather straightforward, but a linear concentration–effect relationship is assumed with sigmoidal dose–response curves. The model does not allow calculation of the variation (confidence interval) within each experiment, only between experiments, and cell kill (a negative fraction) cannot be evaluated.

Isobologram

The first isobolograms were designed in the 1950s by Nobel laureates Elion and Hitchings (13), in which, for each given level of toxicity, the dose of one drug is plotted on the x-axis and that of the second drug on the y-axis (Figure 1B), *e.g.* equitoxic doses of the single drugs. When there is additivity, there should be a straight line connecting the plotted IC_{50} values and when the effect of the combination treatment is synergistic, the plotted line falls to the left of this line and when antagonistic, to the right of this line (14). Subsequently Chou and Talalay (15), computerized the model applying the CalcuSyn program (Biosoft, Cambridge, UK) (see below). Despite the simplicity and accuracy of this method, the extent of synergism or antagonism cannot be quantified, nor is it possible to calculate the variation between experiments, in contrast to fractional-effect analysis. Cell kill cannot be evaluated either. However, this model proved very valuable to move the first combinations into the clinic.

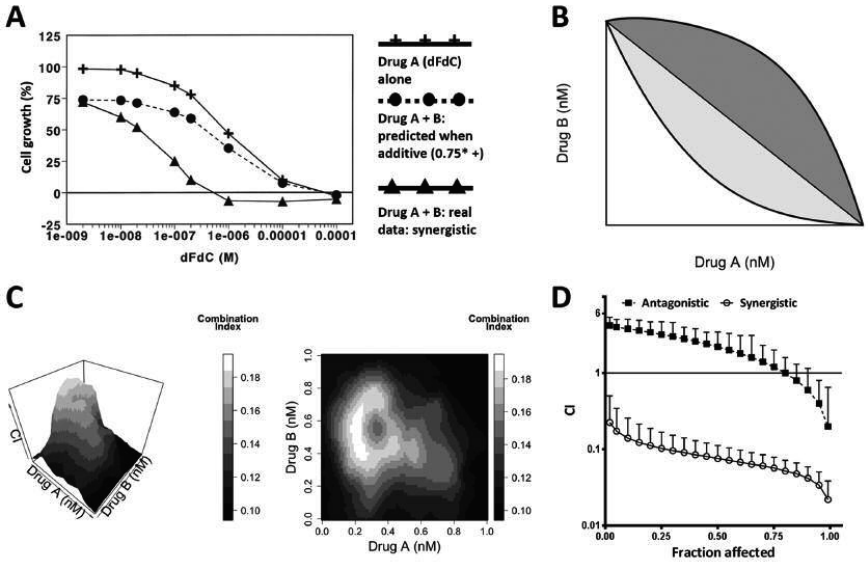


Figure 1. *Examples of plots to determine the effect of combination treatments. A: Fractional-effect analysis, reprinted with permission from (3). B: Isobologram. C: The response surface area model (left) and a contour plot (right). D: Median-effect analysis. CI: Combination index.*

The Response Surface Area Approach

According to several mathematicians (14-16), the response surface area model is to be preferred, since this method allows calculation of the extent of synergism/antagonism (called the envelop of synergism), a confidence interval, and the evaluation of more than two drugs. The results of such analysis are presented in various forms, and usually include a 3-D plot or a contour plot. In the 3-D plot, the concentration of the drugs is plotted on two corresponding horizontal axes and the effect of the combination (response) on the vertical axis (Figure 1C, left). In the initial presentations, synergism was considered a combination index (CI) value >0 and antagonism as less than 0. Later presentations of the 3-D plots used different units for the response (*i.e.* synergy/antagonism), such as a ratio, percentage change or % relative to the control. Synergism is usually presented as a different colour/peak. An alternative presentation is a contour plot, in which the concentration of a drug is on one of the axes and synergism is shown as a 'contour' of a different colour, either white in black/white presentation, or red or blue according to the preference of the mathematician (Figure 1C, right). Despite ready-to-use mathematical models, evaluation of response surface area plots requires a considerable insight into mathematics. Moreover, the presentation of the plots does not allow for determination of whether synergism is observed under clinically relevant conditions, since the extent of growth inhibition is often not shown and also cell kill is often not considered in the calculation. However, in more recent applications, the so-called Bliss index (a measure for synergy) can be given for each effect of a combination. However, in order to optimally predict the efficacy of the combination, experience with the model and sufficient statistical knowledge is essential.

Median Drug-effect Analysis

In order to provide a pragmatic and easy-to-use method to evaluate combinations, Chou and Talalay (15) developed the median-effect principle based on the first models of Loewe (10) and Bliss (11). However, the median-effect principle is based on the mass-action law and not on statistics, and provides a diagnostic plot, including an isobologram. The median drug-effect equation is based on the four equations of Henderson–Hasselbach (19), Michaelis–Menten (20), Hill (21) and Scatchard (22), and therefore the dose–response curves of the single drug and combinations should be sigmoidal. Curves are plotted based on the doses of a drug (D1 and D2), and the effect of a drug, expressed as fraction affected (Fa) and fraction unaffected (Fu). An Fa of 0.5 means 50% growth inhibition, and Fa of 0.75 means 75%

growth inhibition (or 25% growth); this means that Fa+Fu=1. The effect of a drug can be described as the ratio between $Fa/Fu=(D/D_m)^m$. Log transformation yields the median-effect equation: $\log(Fa/Fu)=m \log(D) - m \log(D_m)$, in which D is the dose, m the sigmoidicity of the curve and D_m equals the dose that results in an Fa of 0.5 (*i.e.* IC_{50}). By transforming Fa+Fu=1 to Fu=(1−Fa), the formula can be transformed into $Fa/(1-Fa)=(D/D_m)^m$, and assuming m=1 to $Fa=[1+(D_m/D)]^{-1}$. When the effects for two drugs are combined and the formula for the CI is derived: CI=[(D)1/(D1-Fa)1]+[(D)2/(D1-Fa)2]+[α(D)1(D)2/(D1-Fa)1(D1-Fa)2], in which (D)1 and (D)2 are doses used for combination, (D1−Fa)1 and (D1−Fa)2 are the doses of the individual drugs resulting in 1−Fa; the slope is m, while α=1 for mutually nonexclusive drugs. A CI above 1.2 is considered to be antagonistic, between 0.8 and 1.2 additive, and below 0.8 synergistic.

The usual presentation of the results requires a normal growth-inhibition curve and a plot of CI *versus* Fa (Figure 1D). The growth-inhibition curves are essential for proper evaluation, since it should be determined whether the curves are sigmoidal or hyperbolic, whether there are outliers, and whether sigmoidicity is similar for each drug. A simple quality check includes the comparison of the D_m value calculated by the program with the IC_{50} value from the curve, which should be similar. The CI–Fa plot evaluates synergism over the whole Fa range from 0-1. However, CI values above an Fa of 0.95 are usually not reliable, while CI values obtained at Fa<0.5 can be considered as not relevant, since Fa<0.5 represent minor, clinically not relevant growth inhibition. In order to combine the data of more experiments, it is recommended to calculate the average of the CI values at Fa of 0.5, 0.75 and 0.90 for each separate experiment. Subsequently, the averages of each experiment can be used to calculate the means and SEM for the experiments. In this model, it should be specified whether the combined drugs act on a common target or have a similar mechanism of action, *i.e.* are mutually exclusive, where one agent may prevent the action of the other. The agents may have different mechanisms of action or targets and therefore act in a mutually nonexclusive manner.

In order to digitalize the median-effect method, the CalcuSyn software was designed based on the formulas given above by Chou and Talalay, allowing the user to plot the dose–response curves of the single agents and the combination treatment to determine the CI. CalcuSyn can be downloaded from http://www.biosoft.com/w/calcusyn.htm, however a fee dependent on the intended number of users is required. In 2005, Chou published another software program based on the median-effect of the mass-action law, CompuSyn, which can be downloaded free of charge from www.combosyn.com. The program does not allow values above 1 or below 0 to be entered, which can be solved by plotting these values all at either 0.95 or 0.05, respectively. Although this limitation does not allow evaluation of cell kill, this can be solved by using a modified calculation of Fa as described by Bijnsdorp *et al.* (8). Even though the program advises combining drugs at a fixed ratio based on the IC_{50}, the current version also allows evaluation of non-fixed ratios. Loewe synergy using surface-area models can be calculated using Combenefit, which can be downloaded for free from https://sourceforge.net/projects/combenefit/.

Models evaluating the effect of two agents combined are valuable in the development of novel treatment strategies. Nevertheless, researchers should be aware of the limitations and pitfalls that these models carry. A concise summary of the advantages and disadvantages of the aforementioned methods are listed in Table I.

How Should Novel Tyrosine Kinase-directed Drugs Be Combined?

The models described above were developed primarily to evaluate combinations of two or more conventional anticancer chemotherapeutics. This raises the question of whether the model can also be used to evaluate combinations of conventional chemotherapy agents with novel chemotherapeutics drugs, such as tyrosine kinase inhibitors (TKI) or combinations of TKIs. There is sufficient evidence in literature that Calcusyn can be used to evaluate combinations of drugs such as pemetrexed or gemcitabine with a TKI, such as erlotinib or crizotinib, respectively (23, 24). However, for these studies, it is even more evident that genetic properties of the cells should be taken into account, while the concentration of a TKI should be chosen in a range high enough to modulate its target. For instance, in the case of erlotinib, it is not sensible to combine these drugs for tumours that have a *RAS* mutation (23), while for crizotinib, either a MET proto-oncogene receptor tyrosine kinase (*c-MET*) amplification or mutation is required to achieve an effect (24). This also means that the drugs can better be combined at a non-fixed ratio, for which the current version of CalcuSyn has a separate option.

Similarly, combinations of one or more TKIs can be evaluated using the above-mentioned programs on the condition that genetic and biochemical properties are taken into account when combining these drugs. An example is the combination of erlotinib, an epidermal growth factor receptor (EGFR) inhibitor, with crizotinib, which inhibits c-MET. Resistance to erlotinib can be due to increased c-MET signalling, which can be inhibited by crizotinib. This combination appeared to be synergistic in cells with active EGFR and c-Met signalling, and additive when one of the pathways was active. Antagonism was observed in a cell line lacking these properties (25).

These considerations indicate that a lower concentration of one drug may be sufficient for modulating cellular signalling.

Table I. *Methods for determination of combination effects and their advantages and disadvantages.*

Method	Advantages	Disadvantages
Fractional-effect analysis	Straightforward method for mutually non-exclusive drugs	Model assumes linear concentration-effect curve in contrast to reality (sigmoidal) No confidence interval quantification
Isobologram	Simple and accurate method	Level of synergy or antagonism cannot be determined The isobole based on the dose pair is often curvilinear instead of linear due to variability in drug potency No experimental variation quantification
The response surface approach	Model assumes sigmoidal concentration–effect relationship	Output interpretation requires expertise since it is aggregated Sufficient statistical knowledge essential
Median drug-effect analysis	A distinction between mutually exclusive and non-exclusive acting agents can be made	Synergy evaluation can be complicated Fa values <0 or >1 cannot be included in calculations, but by adapting the model this can still be achieved (8)

Fa: Fraction affected.

This has led some investigators to use ultra-low concentrations in the 'homeopathic' range, arguing that patient toxicity might also be reduced. However, when no clear effect on signalling is observed in mechanistic experiments, such combinations should be considered with care.

Translation of Combinations to *In Vivo* Models

When combining drugs *in vivo*, additional aspects should be taken care of. Ideally a synergistic effect on the tumour would be associated with an antagonistic effect on normal tissues. Unfortunately, this is not often observed and initial dosing *in vivo* (usually mice) should be done with care; moreover, the outcome often cannot be predicted with current *in vitro* alternative models, including organoids, since the whole *in vivo* mechanism of drug absorption, distribution, metabolism and excretion is not present therein. Usually one of these aspects can be investigated in a proper model (*e.g.* cytochrome p450 (CYP)-mediated metabolism or glucuronidation and other phase I and II metabolic pathways or ATP-binding cassette transporter-mediated efflux) and can be used to adapt the *in vivo* scheduling. Therefore, it is advised to start with a lower dose of at least one of the drugs. Usually it is advised to give one drug at the maximal tolerated dose (MTD) of the single agent and the other at its 0.66×MTD, although most investigators start with each drug at its 0.66×MTD. Serial two-fold dilutions can be tested before an efficacy experiment in case severe toxicity is expected. *In vitro* experiments can provide guidelines on the scheduling for animal experiments, but toxicity might also be synergistic, necessitating alternative scheduling. Mechanistic pharmacological studies in animals will provide essential information on proper dosing and scheduling *in vivo*. The above-mentioned example of gemcitabine and cisplatin showed pre-treatment of cell lines with gemcitabine would increase the formation of DNA–platinum adducts and prevent repair of these adducts. However, dosing for both drugs had to be reduced, but the combined effect was superior to that of each drug alone at its MTD (26). The same principle was applied to the combination of gemcitabine and crizotinib in mice, in which the combination was based on *in vitro* synergism in a pancreatic cancer cell line with a *c-MET* amplification. It was decided to give gemcitabine at its 0.83×MTD. The combination of gemcitabine and crizotinib was superior to that of each drug separately, while crizotinib increased the accumulation of the active metabolite of gemcitabine by inhibition of gemcitabine degradation (24). In short, it can be concluded that for *in vivo* combinations, the schedule can often be deduced from *in vitro* data, while dosing should be reduced slightly.

New Computational-based Approaches for Capturing Cellular and Signalling Complexity

The rapid development and application of machine-learning methods is now also being applied in biomedical sciences. The utilisation of several machine-learning methods for assessment of combination drug therapy for HIV, hypertension, infectious diseases and cancer is described elaborately by Tsigelny (27). For example, machine-learning networks can implement various parameters, such as i) compound-specific physical and chemical properties; ii) biochemical response of target molecule(s); iii) cellular response; and iv) patient characteristics, including genomic, proteomic and metabolomic profiles, which sounds promising.

Current predictive models for drug-combination effects are commonly based on high-throughput testing of drug combinations for each cell line. These data are then used to identify the molecular features that predict therapy response (28). A pan-cancer DREAM community challenge, the

AstraZeneca Drug Combination Prediction Challenge, showed such findings on a pan-cancer scale (29). Synergy prediction based on models of drug interactions are appearing.

Given that most therapy combination approaches are based on an aggregated index of synergy, the analytical methods that exist today might need additional refinements to capture intercellular (cell identity) as well as intracellular heterogeneity (signalling activity) and their combined effects on efficacy of drug combinations. Single-cell tracing methods, such as single-cell genetic/mRNA profiling, fluorescent reporter systems and cell-tracing barcode technologies can provide relevant insight into population changes as well as fluctuations in the mechanism of action that result in therapy efficacy. These computational models should provide sufficient complexity to predict the effect of combination therapies on a cellular/signalling level accurately. Given this high complexity, more advanced machine-learning methods, such as deep learning, might be needed to enable adequate modelling.

Conclusion

Drug combinations have been used for decades in the clinic to enhance the treatment of patients, and are therefore not a novelty in the field of pharmacology. Various *in vitro* models together with mathematical equations based on the Bliss independence model, Loewe additivity model or a combination of both, *i.e.* the median drug effect, enable the prediction of the effect of two agents combined. Nonetheless, application of these models should be tailored to the context of each study and with awareness of the limitations and advantages of each method. Considering the increment in the development of novel therapeutics, the number of combinations that can be made is substantial. Aside from more complex single-cell assay read-outs that are able to capture cellular and signalling heterogeneity, more advanced computational models might be needed, including deep-learning methodologies. Therefore, we anticipate that refinements of the classical synergy models with emerging artificial intelligence-based models will benefit the investigation of new combination treatments in the near future.

Conflicts of Interest

The Authors have no conflicts of interest to declare in regard to this study.

Authors' Contributions

BEH, GJ, BW and EG wrote the article. GM provided the figures and GLP the table. MJ, LB, HNW, AVM and MCMK revised the article.

Acknowledgements

This work was partially supported by the following grants: CCA Foundation 2012 and 2015 grants (GJP, EG), Fondazione Pisana per la Scienza (EG), KWF Dutch Cancer Society grants (KWF project#10401 and #11957, EG) and AIRC/Start-Up grant (EG).

References

1 DeVita VT and Chu E: A history of cancer chemotherapy. Cancer Res *68(21)*: 8643-8653, 2008. PMID: 18974103. DOI: 10.1158/0008-5472.CAN-07-6611

2 Bertino JR and Chou T-C: Chemotherapy: synergism and antagonism. *In*: Encyclopedia of Cancer. Vol 1. Academic Press, Inc., pp. 368-379, 1997.

3 Peters GJ, Van Der Wilt CL, Van Moorsel CJA, Kroep JR, Bergman AM and Ackland SP: Basis for effective combination cancer chemotherapy with antimetabolites. Pharmacol Ther *87(2-3)*: 227-253, 2000. PMID: 11008002. DOI: 10.1016/S0163-7258(00)00086-3.

4 Bergman AM, Ruiz Van Haperen VWT, Veerman G, Kuiper CM and Peters GJ: Synergistic interaction between cisplatin and gemcitabine *in vitro*. Clin Cancer Res *2(3)*: 521-530, 1996. PMID: 9816199. DOI: 10.1016/j.canlet.2011.12.016

5 Peters GJ, Backus HH, Freemantle S, Van Triest B, Codacci-Pisanelli G, Van der Wilt CL, Smid K, Lunec J, Calvert AH, Marsh S, McLeod HL, Bloemena E, Meijer S, Jansen G, Van Groeningen CJ and Pinedo HM: Induction of thymidylate synthase as a 5-fluorouracil resistance mechanism. Biochim Biophys Acta Mol Basis Dis *1587(2-3)*: 194-2015, 2002. PMID: 12084461. DOI: 10.1016/S0925-4439(02)00082-0

6 Bergman AM, Pinedo HM and Peters GJ: Steroids affect collateral sensitivity to gemcitabine of multidrug-resistant human lung cancer cells. Eur J Pharmacol *416(1-2)*: 19-24, 2001. PMID: 11282108. DOI: 10.1016/S0014-2999(01)00858-5

7 Perego P, Hempel G, Linder S, Bradshaw TD, Larsen AK, Peters GJ and Phillips RM: Cellular pharmacology studies of anticancer agents: recommendations from the EORTC-PAMM group. Cancer Chemother Pharmacol *81(3)*: 427-441, 2018. PMID: 29285635. DOI: 10.1007/s00280-017-3502-7.

8 Bijnsdorp IV, Giovannetti E and Peters GJ: Analysis of drug interactions. Methods Mol Biol *731*: 421-34, 2011. PMID: 21516426. DOI: 10.1007/978-1-61779-080-5_34

9 Peters GJ, Chatelut E, Larsen AK and Zaffaroni N: EORTC-related new drug discovery and development activities: Role of the Pharmacology and Molecular Mechanisms Group. Eur J Cancer Suppl *10(1)*: 128-140, 2012. DOI: 10.1016/S1359-6349(12)70022-8

10 Loewe S and Muischnek H: Effect of combinations: mathematical basis of problem. Arch Exp Pathol Pharmacol *114*: 313-326, 1926.

11 Bliss CI: The toxicity of poisons applied jointly. Ann Appl Biol *26*: 585-615, 1939. DOI: 10.1111/j.1744-7348.1939.tb06990.x

12 Webb J: Effect of more than one inhibitor. Enzymes Metabol Inh *1*: 66-79, 1963.

13 Elion GB, Singer S and Hitchings GH: Antagonists of nucleic acid derivatives. VIII. Synergism in combinations of biochemically related antimetabolites. J Biol Chem *208(2)*: 477-488, 1954. PMID: 13174557.

14 Steel GG and Peckham MJ: Exploitable mechanisms in combined radiotherapy-chemotherapy: The concept of additivity.

Int J Radiat Oncol Biol Phys *5(1)*: 85-91, 1979. PMID: 422420. DOI: 10.1016/0360-3016(79)90044-0

15 Chou T and Talalay P: Quantitative dose-effect relationships: The combined effects of multiple. Adv Enzyme Regul *22*: 27-55, 1984. PMID: 6382953. DOI: 10.1016/0065-2571(84)90007-4

16 Berenbaum MC: What is synergy? Pharmacol Rev *41(2)*: 93-141, 1989. PMID: 2692037.

17 Di Veroli GY, Fornari C, Wang D, Mollard S, Bramhall JL, Richards FM and Jodrell DI: Combenefit: An interactive platform for the analysis and visualization of drug combinations. Bioinformatics *32(18)*: 2866-2868, 2016. PMID: 27153664. DOI: 10.1093/bioinformatics/btw230

18 Greco WR, Faessel H and Levasseur L: The search for cytotoxic synergy between anticancer agents: A case of Dorothy and the ruby slippers? J Natl Cancer Inst *88(11)*: 699-700, 1996. PMID: 8637018. DOI: 10.1093/jnci/88.11.699

19 Clark WM: The Determination of Hydrogen Ions. Third edition. Baltimore, Williams and Wilkins, 1928.

20 Michaelis L and Menten ML: Die Kinetik der Invertinwirkung. Biochem Z *49*: 333-369, 1913.

21 Hill AV: The Combinations of haemoglobin with oxygen and with carbon monoxide. I Biochem J *7*: 471-480, 1913.

22 Scatchard G and George: The attractions of proteins for small molecules and ions. Ann N Y Acad Sci *51*: 660-672, 1949.

23 Giovannetti E, Lemos C, Tekle C, Smid K, Nannizzi S, Rodriguez JA, Ricciardi S, Danesi R, Giaccone G and Peters GJ: Molecular mechanisms underlying the synergistic interaction of erlotinib, an epidermal growth factor receptor tyrosine kinase inhibitor, with the multitargeted antifolate pemetrexed in non-small-cell lung cancer cells. Mol Pharmacol *73(4)*: 1290-1300, 2008. PMID: 18187583. DOI: 10.1124/mol.107.042382

24 Avan A, Caretti V, Funel N, Galvani E, Maftouh M, Honeywell RJ, Lagerweij T, Van Tellingen O, Campani D, Fuchs D, Verheul HM, Schuurhuis GJ, Boggi U, Peters GJ, Würdinger T and Giovannetti E: Crizotinib inhibits metabolic inactivation of gemcitabine in c-Met-driven pancreatic carcinoma. Cancer Res *73(22)*: 6745-6756, 2013. PMID: 24085787. DOI: 10.1158/0008-5472.CAN-13-0837

25 Van Der Steen N, Leonetti A, Keller K, Dekker H, Funel N, Lardon F, Ruijtenbeek R, Tiseo M, Rolfo C, Pauwels P, Peters GJ and Giovannetti E: Decrease in phospho-PRAS40 plays a role in the synergy between erlotinib and crizotinib in an EGFR and cMET wild-type squamous non-small cell lung cancer cell line. Biochem Pharmacol *166*: 128-138, 2019. PMID: 31078602. DOI: 10.1016/j.bcp.2019.05.014

26 Van Moorsel CJA, Pinedo HM, Smid K, Comijn EM, Voorn DA, Veerman G, Lakerveld B, Van der Vijgh WJF, Giaccone G, Postmus PE and Peters GJ: Schedule-dependent pharmacodynamic effects of gemcitabine and cisplatin in mice bearing Lewis lung murine non-small cell lung tumours. Eur J Cancer *36(18)*: 2420-2429, 2000. PMID: 11094319. DOI: 10.1016/S0959-8049(00)00345-2

27 Tsigelny IF: Artificial intelligence in drug combination therapy. Brief Bioinform, 2018. PMID: 29438494. DOI: 10.1093/bib/bby004

28 Bansal M, Yang J, Karan C, Menden MP, Costello JC, Tang H, Xiao G, Li Y, Allen J, Zhong R, Chen B, Kim M, Wang T, Heiser LM, Realubit R, Mattioli M, Alvarez MJ, Shen Y, Gallahan D, Singer D, Saez-Rodriguez J, Xie Y, Stolovitzky G, Califano A, Abbuehl JP, Altman RB, Balcome S, Bell A, Bender A, Berger B, Bernard J, Bieberich AA, Borboudakis G, Chan C, Chen TH, Choi J, Coelho LP, Creighton CJ, Dampier W, Davisson VJ, Deshpande R, Diao L, Di Camillo B, Dundar M, Ertel A, Goswami CP, Gottlieb A, Gould MN, Goya J, Grau M, Gray JW, Hejase HA, Hoffmann MF, Homicsko K, Homilius M, Hwang W, Ijzerman AP, Kallioniemi O, Karacali B, Kaski S, Kim J, Krishnan A, Lee J, Lee YS, Lenselink EB, Lenz P, Li L, Li J, Liang H, Mpindi JP, Myers CL, Newton MA, Overington JP, Parkkinen J, Prill RJ, Peng J, Pestell R, Qiu P, Rajwa B, Sadanandam A, Sambo F, Sridhar A, Sun W, Toffolo GM, Tozeren A, Troyanskaya OG, Tsamardinos I, Van Vlijmen HWT, Wang W, Wegner JK, Wennerberg K, Van Westen GJP, Xia T, Yang Y, Yao V, Yuan Y, Zeng H, Zhang S, Zhao J and Zhou J: A community computational challenge to predict the activity of pairs of compounds. Nat Biotechnol *32(12)*: 1213-1222, 2014. PMID: 25419740. DOI: 10.1038/nbt.3052

29 Menden MP, Casale FP, Stephan J, Bignell GR, Iorio F, McDermott U, Garnett MJ, Saez-Rodriguez J and Stegle O: The germline genetic component of drug sensitivity in cancer cell lines. Nat Commun *9(1)*: 3385, 2018. PMID: 30139972. DOI: 10.1038/s41467-018-05811-3

Received May 14, 2019
Revised June 7, 2019
Accepted June 12, 2019

Chapter 9

"Open Sesame?": Biomarker Status of the Human Equilibrative Nucleoside Transporter-1 and Molecular Mechanisms Influencing its Expression and Activity in the Uptake and Cytotoxicity of Gemcitabine in Pancreatic Cancer

Randazzo O*, Papini F*, Mantini G, Gregori A, Parrino B, Liu DSK, Cascioferro S, Carbone D, Peters GJ, Frampton AE**, Garajova I*, Giovannetti E**
*These authors contributed equally
** These authors contributed equally

Cancers, 2020

cancers

MDPI

Article

"Open Sesame?": Biomarker Status of the Human Equilibrative Nucleoside Transporter-1 and Molecular Mechanisms Influencing its Expression and Activity in the Uptake and Cytotoxicity of Gemcitabine in Pancreatic Cancer

Ornella Randazzo [1,2,†], Filippo Papini [1,†], Giulia Mantini [1,3], Alessandro Gregori [1], Barbara Parrino [2], Daniel S. K. Liu [4], Stella Cascioferro [2], Daniela Carbone [2], Godefridus J. Peters [1,5], Adam E. Frampton [4,6,*], Ingrid Garajova [1,7,†] and Elisa Giovannetti [1,3,*]

1 Department of Medical Oncology, Cancer Center Amsterdam, Amsterdam UMC, VU University Medical Center (VUmc), 1081 HV Amsterdam, The Netherlands; o.randazzo@amsterdamumc.nl (O.R.); f.papini@student.vu.nl (F.P.); g.mantini@amsterdamumc.nl (G.M.); a.gregori@amsterdamumc.nl (A.G.); gj.peters@amsterdamumc.nl (G.J.P.); igarajova@ao.pr.it (I.G.)

2 Dipartimento di Scienze e Tecnologie Biologiche Chimiche e Farmaceutiche (STEBICEF), Università degli Studi di Palermo, 90123 Palermo, Italy; barbara.parrino@unipa.it (B.P.); stellamaria.cascioferro@unipa.it (S.C.); daniela.carbone@unipa.it (D.C.)

3 Cancer Pharmacology Lab, AIRC Start Up Unit, Fondazione Pisana per la Scienza, 56017 Pisa, Italy

4 Division of Cancer, Department of Surgery & Cancer, Imperial College, Hammersmith Hospital campus, London W12 0NN, UK;; daniel.liu08@imperial.ac.uk

5 Department of Biochemistry, Medical University of Gdansk, 80-210 Gdansk, Poland

6 Faculty of Health and Medical Sciences, The Leggett Building, University of Surrey, Guildford GU2 7XH, UK

7 Medical Oncology Unit, University Hospital of Parma, Via Gramsci 14, 43126 Parma, Italy

* Correspondence: a.frampton@imperial.ac.uk (A.E.F.); e.giovannetti@amsterdamumc.nl (E.G.); Tel.: +31-003-120-444-2633 (E.G.)

† These authors contributed equally to this paper.

Received: 28 September 2020; Accepted: 26 October 2020; Published: 31 October 2020

check for updates

Simple Summary: Despite the enormous advance in biomarker discovery, many potential biomarkers of drug activity are unable to satisfy the clinical need due to inadequate sensitivity and specificity. The nucleoside transporter hENT-1 has been studied as a potential biomarker to predict the effect of the widely used anticancer drug gemcitabine in pancreatic cancer. However, several studies showed controversial results regarding the predictive value of hENT-1, prompting new analyses with larger cohorts of patients and standardized methodologies. Improved insights on molecular mechanisms underlying hENT-1 expression and activity should also help in the identification of subsets of patients who are more likely to benefit from specific treatments and improve their clinical outcome. The establishment of such biomarker is especially valuable in pancreatic cancer, which is frequently characterized by complex disease biology and high mortality.

Abstract: Pancreatic ductal adenocarcinoma (PDAC) is an extremely aggressive tumor characterized by early invasiveness, rapid progression and resistance to treatment. For more than twenty years, gemcitabine has been the main therapy for PDAC both in the palliative and adjuvant setting. After the introduction of FOLFIRINOX as an upfront treatment for metastatic disease, gemcitabine is still commonly used in combination with nab-paclitaxel as an alternative first-line regimen, as well as a monotherapy in elderly patients unfit for combination chemotherapy. As a hydrophilic nucleoside analogue, gemcitabine requires nucleoside transporters to permeate the plasma membrane, and a major role in the uptake of this drug is played by human equilibrative nucleoside transporter 1

Cancers **2020**, *12*, 3206

(hENT-1). Several studies have proposed hENT-1 as a biomarker for gemcitabine efficacy in PDAC. A recent comprehensive multimodal analysis of hENT-1 status evaluated its predictive role by both immunohistochemistry (with five different antibodies), and quantitative-PCR, supporting the use of the 10D7G2 antibody. High hENT-1 levels observed with this antibody were associated with prolonged disease-free status and overall-survival in patients receiving gemcitabine adjuvant chemotherapy. This commentary aims to critically discuss this analysis and lists molecular factors influencing hENT-1 expression. Improved knowledge on these factors should help the identification of subgroups of patients who may benefit from specific therapies and overcome the limitations of traditional biomarker studies.

Keywords: pancreatic cancer; drug resistance; human equilibrative nucleoside transporter 1; clinical outcome

1. Introduction

In the story of "Ali Baba and the Forty Thieves" from the book *One Thousand and One Nights*, "Open sesame" is the magical phrase that opens the mouth of a cave in which the thieves have hidden a treasure. This statement has been commonly used to define something that allows a person to do or enter something easily, or something that unfailingly brings about a desired end. Several cellular transporters are essential for the entry (or efflux) of anticancer drugs [1,2] but, despite a number of preclinical and clinical studies on their role as biomarkers and targets, they have not yet been exploited or exploited correctly to improve clinical outcome.

The human equilibrative nucleoside transporter (hENT-1) represents a quintessential example of such a transporter. This protein is indeed the main transporter involved in the entrance of nucleoside analogs and has attracted extensive attention for its potential role as predictive biomarker for the anticancer activity of gemcitabine [3], as well as for the development of drugs bypassing this transporter in order to overcome gemcitabine resistance in pancreatic ductal adenocarcinoma (PDAC) [4–6]. In particular, we have read with great interest the recent comprehensive multimodal analysis of hENT-1 status, which has been performed by Raffenne and collaborators in the largest cohort of PDAC patients to date [3]. In the present commentary, we summarize the key findings of this analysis and discuss further insights on molecular and pharmacological factors influencing the role of nucleoside transporters in the uptake and cytotoxicity of gemcitabine in PDAC.

2. Pancreatic Ductal Adenocarcinoma (PDAC)

Pancreatic ductal adenocarcinoma (PDAC) is the most common form of pancreatic cancer and is amongst the deadliest solid malignancies [7]. Despite extensive genetic mapping elucidating key mechanisms in PDAC initiation and progression [8], both conventional and experimental drugs showed limited effects [7].

PDAC is indeed a highly invasive and aggressive disease with an overall 5-year survival rate lower than 10% [7]. Several factors are responsible for this grim prognosis, including late diagnosis and lack of effective therapies [9,10]. Surgery represents the only curative intervention available for patients with local disease. However, even after successful tumor resection and adjuvant chemotherapy, the 5-year survival rate is around 25–30%, with most of the patients experiencing tumor recurrence and metastatic disease within 6–24 months from surgery [11,12]. Chemotherapy is the only treatment available for patients with advanced or metastatic disease, which includes the vast majority of diagnosed PDAC cases [7].

Unfortunately, PDAC is an inherently resistant disease, with low percentages of response rate to all current treatment regimens. Moreover, even when chemotherapy is initially effective, chemoresistance typically occurs after a few cycles, leading to disease progression and mortality [13,14]. This resistance is caused by both cellular intrinsic and extrinsic factors such

as cancer stem cells (CSCs), activation of the epithelial-mesenchymal transition (EMT), and presence of a highly desmoplastic and immunosuppressive tumor microenvironment (TME) [15–18]. Understanding these molecular mechanisms of chemoresistance is an essential step towards increasing the efficacy of treatments and clinical outcome.

Despite the development of more effective multi-drug regimens such as FOLFIRINOX (a chemotherapy regimen made up of the following four drugs: FOL—folinic acid, F—fluorouracil (5-FU), IRIN—irinotecan, OX—oxaliplatin), gemcitabine is still widely used for PDAC treatment, both combined with nab-paclitaxel (Abraxane®, Celgene, Summit, NJ, USA), and as monotherapy, especially in patients who are unfit for more toxic poly-chemotherapy regimens [14]. Several studies have focused on molecular intracellular determinants of gemcitabine activity and metabolism. Among these molecular determinants, mRNA and protein expression of the equilibrative transporter-1 (hENT-1) emerged as potential predictors of drug activity in a number of preclinical and clinical studies in PDAC.

3. Nucleoside Transporters Involved in Gemcitabine Uptake

The two major classes of nucleoside transporters that have been described in mammalian cells include the concentrative nucleoside transporters (CNTs) and the equilibrative nucleoside transporters (ENTs). These transporters are transmembrane glycoproteins that localize to the cellular and mitochondrial membranes, but can also be found in lysosomes [19] and mediate the cellular uptake of nucleosides required for nucleotide synthesis in cells that lack de novo nucleotide synthesis pathways.

CNTs mediate the inward Na^+-dependent transport whereas ENTs are bi-directional facilitators of the transmembrane flux of nucleosides [20]. ENTs can be found in almost all cell types unlike CNTs, which are present in intestinal and renal epithelia [21], as well as in hepatocyte cells and in chorionic villi of human term placenta [22].

As bi-directional carriers, ENTs regulate both the influx and efflux of substrates. The human ENT homologues (hENTs) are classified into four groups: hENT-1 (SLC29A1), hENT-2 (SLC29A2), hENT-3 (SLC29A3) and hENT-4 (SLC29A4) [23]. The hENT-1 is sensitive to nitrobenzylmercaptopurine ribonucleoside (NBMPR) to which it binds with a high affinity [24]; while hENT-2 is insensitive to inhibition by NBMPR. However, the hENT1–3 shows selectivity to the NBMPR substrate which also blocks hENT-4 albeit at higher concentration than required for hENT-1 [25]. Of note, hENT-4 is better known as the plasma membrane monoamine transporter because it carries organic cations such as biogenic amines and neurotoxins. This transporter mediates the transport of adenosine in a pH-dependent manner and its activity increases in acidic conditions (optimal transport at pH 6.0) [26].

These transporters are involved in the uptake of several drugs (Figure 1) with different chemical structures (Figure S1).

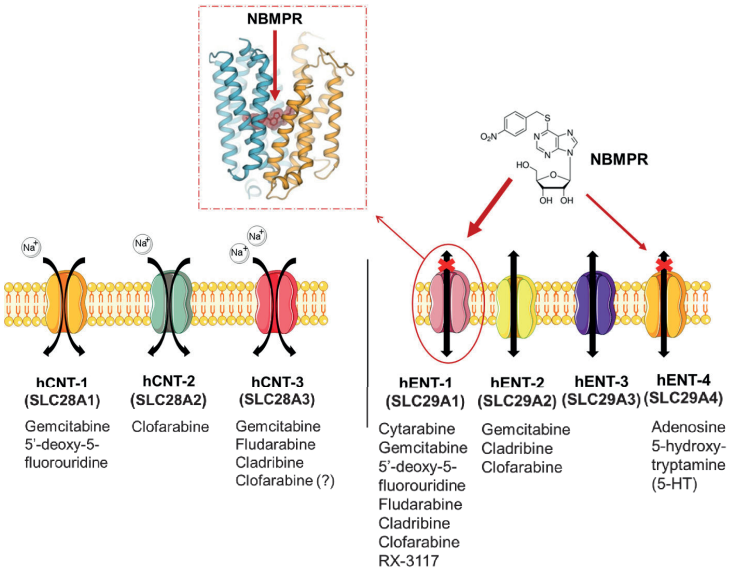


Figure 1. Nucleoside transporters in mammalian cells. CNTs facilitate the Na^+-dependent transport while ENTs are bi-directional and Na^+-independent transporters, facilitating the uptake of different anticancer drugs [27–30]. Of note, hENT-1 (es) is NBMPR-sensitive and the figure shows its crystalized structure [30]. Acronyms: 5-HT, 5-hydroxy-tryptamine; NBMPR, nitrobenzylmercaptopurine ribonucleoside.

Gemcitabine (2′,2′-difluorodeoxycytidine, dFdC) is a pyrimidine analog and relies on membrane transporters for its intracellular uptake [31,32]. The uptake of gemcitabine can be mediated by hENT-1/2 and hCNT-1/3. However, hENT-1 and hCNT-1 appear to be the most efficient transporters involved in the entry of gemcitabine into cells. Notably, hENT-1, the most widely expressed nucleoside transporter in human tissues, is overexpressed in different tumor types, including PDAC (Figure S2A). This is reported in the GEPIA web server, analyzing the RNA sequencing expression data of 9736 tumors and 8587 normal samples from the TCGA and the GTEx projects [33]. However, the comparison of TCGA-PAAD data of pancreatic cancer specimens with the matched non-tumor samples as well as the analysis of similarly matched transcriptomics and proteomics public datasets did not show a significance difference in hENT-1 expression levels (Figure S2B).

Structurally, hENT-1 is a 456-residue protein (50 kDa) with 11 transmembrane domains and three N-linked glycosylation sites, which are essential for localization, function and oligomerization. The first glycosylation site (Asn 48) is located between the first and second transmembrane domains in the hydrophilic loop, whereas the other two sites (Asn 277 and 288) are between the sixth and seventh transmembrane domains [32]. The Km for purine and pyrimidine nucleosides transport range from 0.05 mM to 0.60 mM according to a study performed in *Xenopus laevis* [34].

4. Role of hENT-1 in Gemcitabine Activity as Potential Predictive Biomarker

Several studies showed that hENT-1 expression is essential for gemcitabine cytotoxic effects [35,36]. Higher expression levels of hENT-1 have indeed been associated with higher uptake and activity of gemcitabine in cancer cells, using different preclinical models [37–40].

A number of retrospective studies on hENT-1 mRNA and protein expression with PCR and immunohistochemical (IHC) methodology demonstrated that high levels of hENT-1 correlated with a statistically significant longer survival (Figure 2), both in the adjuvant and in the metastatic

setting, though the number of patients in the latter cohort was extremely small [41–52]. For instance, a retrospective analysis of a cohort of PDAC patients from the RTOG9704 phase III clinical trial, which compared gemcitabine with 5-FU after surgical resection, showed an association between high tumor hENT-1 expression and increased overall survival (OS) in patients who received gemcitabine (n = 91), but not in those who received 5-FU [43]. These data support the role of hENT-1 as a specific predictive biomarker for the efficacy of gemcitabine.

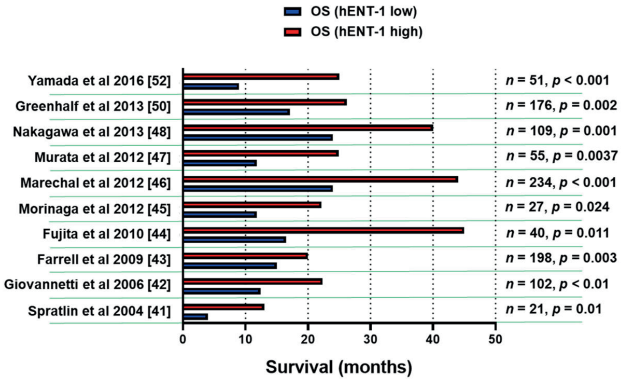


Figure 2. Association of hENT-1 expression levels and survival. Bar graph summarizing the survival data of PDAC patients in studies supporting the role of hENT-1 as a predictive biomarker for the efficacy of gemcitabine. The clinicopathological and treatment details of studies before 2014 [41–48] were reviewed systematically by Nordh et al. [49], while a more recent systematic review by Bird et al. [51] evaluated also the data from the European Study Group for Pancreatic Cancer 3 (ESPAC-3) trial [50]. n = number of patients, p = statistical p values.

Conversely, high expression levels of hENT-1 emerged as a prognostic biomarker of poor outcome in cholangiocarcinoma. Indeed, while a first study showed a significant association between disease-free survival (DFS) and high expression of membrane hENT-1 in gemcitabine-treated patients [53,54], hENT-1 overexpression was associated to high proliferation rate and significantly shorter survival in resected intrahepatic cholangiocarcinoma patients who did not receive adjuvant treatments [55]. This might be explained by the different levels of hENT-1 expression and proliferative rates in different tumor types and warrants further, larger studies.

Up to now, the largest prospective study on hENT-1 in PDAC has been performed within the European Study Group for Pancreatic Cancer 3 (ESPAC-3) trial. This study highlighted a significant association between high hENT-1 protein expression and longer OS and DFS in PDAC patients receiving gemcitabine-based chemotherapy post-surgery [50]. Moreover, a quantitative metanalysis including 7 studies with a total of 770 patients (405 hENT-1-negative and 365 hENT-1-positive). showed that hENT-1 expression was significantly associated with both prolonged DFS (HR 0.58, 95% CI, 0.42–0.79) and OS (HR 0.52, 95% CI, 0.38–0.72) in patients receiving adjuvant gemcitabine-based therapy [52].

In contrast, Kawada et al. [56], who evaluated hENT-1 expression in PDAC patients undergoing neoadjuvant chemoradiation, showed only a trend towards statistically significant better disease-specific survival in patients with low expression of hENT-1. These results might be explained by the potential preferential eradication of tumor cells with high expression of hENT-1 by the neoadjuvant treatment before the tumor samples collection. This issue can be overcome by the use of fine-needle aspiration biopsy (FNAB) for retrieving cancer cells before treatment and resection of the tumor. Indeed, a study on pretreatment hENT-1 expression in endoscopic ultrasonography-guided FNAB specimens obtained from resectable, borderline-resectable, and locally advanced PDAC, showed that hENT-1 expression was an independent prognostic factor in both whole patients and those with resection [52]. Regardless of T3 and T4, hENT-1-positive patients with resection had significantly better prognosis

than hENT-1-negative patients, whose prognosis was similar to those without resection, suggesting that the evaluation of hENT-1 expression using FNAB samples before chemoradiation provides useful information on patients who might benefit from curative-intent resection.

However, other recent studies on the evaluation of hENT-1 status using IHC in PDAC patients reported conflicting results, possibly due to the use of different antibodies. In particular, the analysis of samples from 156 patients enrolled in the CONKO-001 phase III trial did not show a significant association of high hENT-1 expression with improved median DFS or OS [57]. Similar negative results were observed within a retrospective translational subgroup analysis for hENT-1 in 130 samples from patients enrolled in the AIO-PK0104 multicenter phase III trial [58]. In both cases the researchers used the rabbit monoclonal antibody SP120 and suggested to perform a parallel study using the rabbit 10D7GD antibody. Such study was performed by Marechal and collaborators as well as by Svrcek and collaborators who reported that hENT-1 status was predictive of gemcitabine benefit in patients receiving gemcitabine-adjuvant chemotherapy when evaluated with the 10D7G2 antibody, yet no predictive value was observed when hENT-1 status was assessed using the SP120 antibody [46,59].

In contrast, Kalloger and collaborators, who performed the staining with 10D7G2 and SP120 antibodies (both optimized to run on the Ventana platform), in samples from 227 patients, suggested that both these antibodies can be used to predict gemcitabine sensitivity in resected PDAC [60]. This study suggested also that the use of both antibodies and of the percentage of cells staining positive for hENT-1, instead of the H-score methodology, add critical information that enables the stratification of patients with good, intermediate, or poor response to adjuvant gemcitabine. However, as recognized by the authors: "these findings need to be externally validated in cohorts derived from randomized controlled trials" [60].

Overall, these controversial findings question the predictive value of the available anti-hENT-1 antibodies and call for the establishment of a standardized IHC methodology before hENT-1 status could be used as a predictive biomarker in the clinical setting.

5. Evaluation of the Study "hENT-1 Status in PDAC Patients—Are We Ready Yet?"

In a recent study Raffenne and collaborators provided the most comprehensive multimodal analysis of hENT-1 status in the largest cohort of PDAC patients to date (i.e., 471 patients with resected PDAC) [3]. In this study the expression of hENT-1 evaluated using the 10D7G2 antibody was predictive of both prolonged DFS and OS in PDAC patients receiving gemcitabine adjuvant chemotherapy. In contrast, no predictive value of gemcitabine benefit was observed when hENT-1 status was assessed using the SP120 clone when comparing surgery-gemcitabine vs. surgery-only groups. Three additional antibodies (PAB2255, MC-9777, and 11337-1-AP) manufactured by three different companies (MBL™, Woburn, MA, USA; Abnova™, Taipei, Taiwan; and Acris™-OriGene™, Rockville, MD, USA) were further evaluated to establish their potential for the analysis of hENT-1 status. None of these antibodies showed a predictive value of gemcitabine benefit over controls, providing compelling evidence that commercially available anti-hENT-1 antibodies are not suitable for the evaluation of hENT-1 status. Interestingly, all tested antibodies, except the 10D7G2 clone, detected multiple bands on Western blot that did not correspond to the expected glycosylated forms of hENT-1, hence suggesting that these antibodies may recognize and bind to non-completely functional forms of the hENT-1 protein.

Raffenne and collaborators also evaluated the predictive value of mRNA expression levels of hENT-1 which was assessed using microarray data and qRT-PCR analyses performed on formalin-fixed paraffin-embedded (FFPE) tissues from resected specimens. No difference in both DFS and OS was observed when the median hENT-1 mRNA value was used to discriminate between hENT-1 high- and low-expressing tumors. Nonetheless, an increasingly predictive trend was detected when more stringent thresholds were employed (top 25% for OS and top 10% for DFS). Further increase of the threshold (top 10% vs. bottom 10% or bottom 90%) allowed the selection of a population of exceptional gemcitabine responders.

These results might be explained by the fact that the authors used whole tumors. Because of the dense stromal reaction, the analysis of PCR data in PDAC specimens should indeed be performed only after careful evaluation of the percentage of tumor cells and, when feasible, after laser-microdissection [61].

For instance, our PCR analysis of 22 non-microdissected (no LMD, including tumor and stroma tissues) samples showed a minor gene expression variability, with coefficient of variation values of the hENT-1 expression values ranging from 7% to 16% compared to the respective microdissected (LMD, including only tumor tissues) specimens (Figure 3A,B). This could potentially affect the stratification of the patients in the "low" vs. "high" expression categories and the correlation with clinical outcome. Additionally, proteomics analyses of LMD matched epithelial and stromal compartments showed an up-regulation ($p = 0.017$) of hENT-1 in the epithelial compartment (Figure 3C). Of note, although the presented cohort is relatively small ($n = 13$), epithelial hENT-1 expression was associated with significantly longer survival while no difference in the OS curve were observed for hENT-1 stromal expression (Figure 3C). Successful dissection of tumor and stromal compartment is reported in Figure S3. Recent studies showed the impact of LMD on the quality of both mRNA and protein content in PDAC specimens [61,62], and might explain why a not laser-assisted microdissection did not result in the association of hENT-1 expression levels with disease- specific survival, as reported by Jiraskova and collaborators in a retrospective study on a cohort of 69 resected PDAC patients treated with gemcitabine [63].

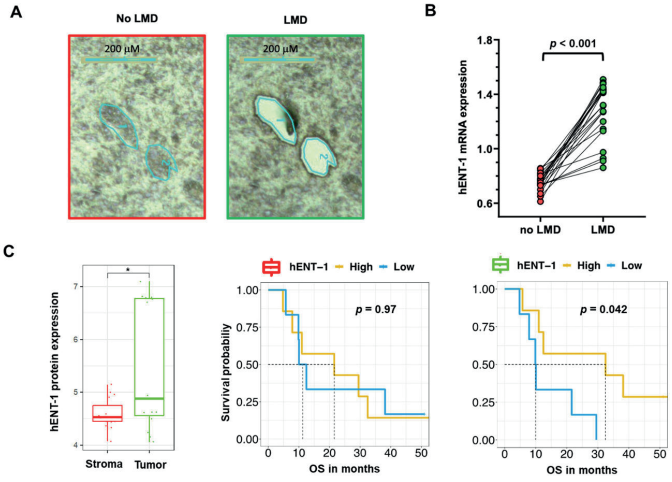


Figure 3. Impact of laser-microdissection on mRNA expression levels of hENT-1 in PDAC samples (**A**) Representative pictures of PDAC epithelium and stroma before and after laser-assisted microdissection, performed on frozen tissue with the Leica LDM7000 microscope, as described previously [62]. (**B**) Comparison between hENT-1 gene expression values in microdissected (LMD) and non-microdissected (no LMD) samples from 22 PDAC tissues. Expression of hENT-1 was evaluated by Real-Time Quantitative PCR and normalized to GAPDH expression, as described previously [42]. *t*-test statistics was performed using Graph Pad version 7. Clinicopathological features of this subgroup of patients are reported in Table S1 [64–66]. (**C**) Comparison between hENT-1 protein expression values (evaluated by nanoLC-MS/MS) in LMD stromal and tumor compartments (* = $p < 0.05$) in $n = 15$ matched samples with respective survival curves. Protein values are represented in log10(LFQ). Patients were grouped in high and low hENT-1 protein expression according to the median cutoff. [62,67]. T-test statistics and Kaplan-Meier plots were performed in R version 3.5.2.

Of note, the data of this study also suggested a limited proportional dependence between hENT-1 gene expression evaluated by qRT-PCR in FFPE samples and protein levels as assessed by immunohistochemistry with the 10D7G2 antibody [63]. This is in agreement with the study by Raffenne and collaborators, where a higher degree of concordance was observed between hENT-1 mRNA expression levels and the SP120 rabbit clone [3]. These findings suggest that this antibody might recognize an unprocessed form of hENT-1 which is directly linked to the mRNA level, whereas the 10D7G2 clone recognizes the active glycosylated stabilized form, hence explaining its better predictive value. Remarkably, a significant correlation between mRNA level and IHC was also found, for another non-commercial rabbit anti-hENT-1 antibody developed by Pastor-Anglada and collaborators, who observed similar results in tumor cells with a different pathology [40,68].

Lastly, in the study by Raffenne and collaborators, the hENT-1 status was assessed by IHC using the 10D7G2 antibody also on coupled samples from both primary and metastatic tumors. Concordance between primary tumor and metastases was excellent for synchronous metastases. In contrast, a lower concordance was reported between metachronous metastases and primary tumors. As postulated by the authors, this discrepancy could be the result of the gemcitabine treatment that led to the selection of hENT-1-low clones in the metachronous metastases [3]. In this regard, further limitations might be represented by the relatively small sample size and other determinants influencing hENT-1 expression, such as disease stage and parameters discussed in Section 6. This is an extremely important aspect that could explain why the role of hENT-1 expression as a biomarker could not be validated when a comparison of gemcitabine with its lipophilic analog CO-101 was carried out within the prospective biomarker-stratified Low hENT1 Adenocarcinoma of the Pancreas (LEAP) trial, which enrolled PDAC patients in the metastatic setting [5]. However, another potential explanation of the lack of association between hENT-1 expression and response to gemcitabine is the use of the rabbit monoclonal antibody SP120, as reported above.

6. Factors Involved in hENT1 Regulation and Gemcitabine Activity

6.1. Genetics: Mutations and Polymorphisms

Structure and function studies have reported that hENT-1 transmembrane domains (TMDs 3–6) might be involved in interaction of nucleosides with the transporter [69]. Based on that, SenGupta et al. [70] explored the role of point mutations on glycine 179 and glycine 184 located in transmembrane domain five (TMD 5), using a GFP-tagged hENT-1 in a yeast nucleoside transporter assay system. As a result, substitution of glycine 179 with leucine, valine, or cysteine caused the lack of transporter activity without affecting its targeting to the plasma membrane. On the other hand, mutation of glycine 179 to alanine or serine influenced neither the activity of hENT-1 nor its targeting to plasma membrane. Hence, it could conceivably be suggested that glycine 179 may have an indirect but vital role in the permeation pathway of hENT-1.

Single nucleotide polymorphisms (SNPs) have been mentioned to affect the gene expression of hENT-1 and thus influencing gemcitabine clinical efficacy [71,72]. Three SNPs were confirmed in the proximal promoter of hENT-1 by Myers et al. [72]: −1345C > G, −1050G > A, and −706G > C. Higher expression levels were observed for two haplotypes (CGC, CAG) when cloning the four naturally occurring haplotypes (CGG, CAG, CGC, GAG) into a Luciferase expression system. Individuals with such haplotypes presented increased hENT-1 expression in comparison to those with normal haplotypes.

The distribution of variants in genes involved in gemcitabine pharmacology and their association with non-small lung cancer was further characterized by Soo et al. [73]. Their results revealed that the non-synonymous variant CNT1 + 1561 G > A is correlated with increased uptake of gemcitabine and hematologic toxicity. However, as the study was limited by the small sample size, larger studies are needed to validate these findings.

6.2. Epigenetics and microRNAs

6.2.1. Epigenetics

The expression of drug transporters, drug metabolizing enzymes, and nuclear receptors, is under epigenetic control affecting the regulation of various genes and response to chemotherapeutic drugs [74]. The cellular levels of three histone modifications (H3K4me2, H3K9me2, H3K18ac), were examined by tissue microarrays from two cohorts with PDAC patients by Manuyakom and collaborators. Low H3K4me2 and H3K9me2 levels were associated with worse overall and disease-free survival (Adjusted HR: 1.48 and 1.44, respectively) [75]. Later, methylation of lysine H3K9 was extensively studied by Candelaria, et al. [76]. More specifically, they exposed CaLo cells to increasing concentrations of gemcitabine which eventually became resistant. This state was accompanied by down regulation of hENT-1. To determine whether gemcitabine resistance was associated with gene silencing induced by increased histone deacetylase activity, they performed ChIP assays, which finally showed a decrease in H3 and H4 acetylation at the hENT-1 promoter. They proposed that this mechanism could silence the expression of hENT-1 and therefore lead to gemcitabine resistance in cervical cancer cell lines.

6.2.2. microRNAs

In the recent years it has become clear that protein expression levels can be regulated by microRNAs. microRNAs are small non-coding RNAs, of 19 to 25 nucleotides, that interact with the mRNA of coding genes, directing their post-translational repression. They are known to influence various cellular processes such as cell proliferation and cell death, mainly through negative regulation of gene expression [77]. In pancreatic cancer, several miRNAs have been reported to be aberrantly expressed, including miR-34 [78], miR-21, miR-155, miR-221, and miR-222 [79]. The regulation of nucleoside transporters by microRNAs is still poorly understood. Theoretically microRNAs could target the mRNA of nucleoside transporters, down-regulating their expression levels.

Using a collection of databases of microRNA-gene interactions ("multimir" R package), 175 miRNAs emerged as miRNA potentially targeting hENT-1 (Table S2). Of note, four of these microRNAs are overexpressed in PDAC as reported in Table 1.

Table 1. MicroRNAs potentially influencing hENT1 expression.

microRNA	microRNA acc	Experiment Type	Database Source	Comments	Reference
hsa-miR-196a-3p	MIMAT0004562	PAR-CLIP	mirtarbase	Up-regulated in exosomes of PDAC's serum	[67]
hsa-miR-221-5p	MIMAT0004568	Degradome sequencing	tarbase	Up-regulated in PDAC cancer stem cells	[70]
hsa-miR-23b-3p	MIMAT0000418	Degradome sequencing	tarbase	Up-regulated in exosomes of PDAC's serum and correlated to CA19–9	[68]
hsa-miR-155-5p	MIMAT0000646	Degradome sequencing	tarbase	Up-regulated in GEM resistant PDAC cells	[69]

Of note, the presence of tumor-derived microRNAs in both tissues and body fluids offers an opportunity for their potential application as liquid biopsy-based biomarkers, and future studies should evaluate whether emerging circulating microRNAs could be a useful tool for minimally-invasive estimation of hENT-1 levels and prediction of gemcitabine activity in PDAC patients.

MiR-196a-3p is up-regulated in exosomes of pancreatic cancer cell lines and in serum's exosomes of localized PDAC patients (stage I and IIA, $n = 15$) when compared to healthy subjects ($n = 15$) [80]. However, data on outcome or to response to gemcitabine are missing.

Conversely, miR-23b-3p was found to correlate with pancreatic cancer progression in a cohort study in patients with chronic pancreatitis (CP) and pancreatic cancer. Furthermore, in this study, the authors did not provide data on gemcitabine activity, but assessed the expression level of miR-23b-3p in exosomes isolated from patients' serum demonstrating the association of miR-23b-3p to CA19–9 levels [81].

High levels of miR-155–5p were associated to gemcitabine resistance in a study conducted by Mikamori and colleagues [82]. They reported three important findings: (i) long-term exposure to gemcitabine resulted in an increasing level of miR-155–5p; (ii) miR-155–5p levels were positively associated to exosome secretion that promoted gemcitabine resistance; (iii) increasing level of miR-155–5p in PDAC cell while blocking exosome secretion did not induce gemcitabine resistance. This later finding suggests an indirect activity which might be mediated by the miRNA mediated modulation of hENT-1 levels and deserves further research.

Similarly, Zhao and colleagues validated the antagomir for miR-221–5p (a group of miRNA antisense oligonucleotides) to restore chemosensitivity in gemcitabine-resistance cell lines. This miRNA was over-expressed in PDAC cancer-stem-cell subpopulation and regulated some stemness markers such as CDK6, C5ORF41, EFNA1, IRAK3, KLF12, MAPK10, NRP1, SMAD7, SOCS6 and ZBTB41 [83]. However, more research needs to be performed to determine the prognostic and/or predictive characteristics of both tissue and circulating microRNAs regarding their role in nucleoside transport.

6.3. Tumour Microenvironment

6.3.1. Hypoxia

PDAC is characterized by a unique desmoplastic stroma and by the presence of an intense fibro-inflammatory reaction, known as desmoplastic reaction (DR). DR causes the continuous deposition of extracellular matrix components, including collagen type I and III, hyaluronic acid, and fibronectin, by activated pancreatic stellate cells (PSCs) [84]. As a result of increased intratumoral pressure and the subsequent compression of tumor vasculature, tumor cells experience hypoxia and metabolic stress [85]. Koong et al. were the first to observe hypoxia in PDAC [86], reporting areas of pancreatic carcinoma tissues with median pO_2 levels of 0–5.3 mmHg. In contrast, normal tissues had a median pO_2 level of 24 to 92.7 mmHg. Later, Buchler and collaborators [87] showed that hypoxia-inducible factor 1 (HIF-1), an important regulator of cellular response to hypoxia, is activated in PDAC in response to low oxygen conditions. High levels of HIF-1α promote angiogenesis via increased VEGF expression [87], hence promoting PDAC proliferation and metastatic potential [88]. Based on this rationale, clinical trials using drugs targeting angiogenesis have been conducted. However, despite promising preliminary results, anti-angiogenic drugs demonstrated low efficacy in PDAC. Low drug delivery due to vasculature collapse and poor tumor perfusion might explain, at least in part, the modest effectiveness of anti-angiogenic drugs (e.g., bevacizumab) in PDAC [89].

Remarkably, hypoxia also influences nucleoside transporters, as described by Eltzschig and collaborators [90]. Using in vitro and in vivo models of extracellular adenosine signalling, it was shown that hENT-1 and hENT-2 gene expression and function are negatively regulated by HIF-1α.

In particular, hENT-1 and hENT-2 are involved in the passage of adenosine through the endothelial membrane acting as bi-directional channels in normoxic conditions. In contrast, under hypoxic conditions the adenosine movement is unidirectional but predominantly inward because the extracellular adenosine concentration is much higher than the intracellular. Therefore, the repression of NTs induced by hypoxia causes an extracellular increase of adenosine concentration and signaling effects. Additionally, HIF-α forms a heterodimer with HIF-β during hypoxia which also causes the nuclear translocation of HIF-1 and the binding to the promoter of hypoxia-responsive element of hENT-1 (Figure 4). Thus, a downregulation of hENT1 occurs and consequently a decrease of adenosine uptake. These mechanisms represent a transcriptional pathway to limit the inflammatory response and to ensure the integrity of the vascular barrier during hypoxic conditions [91].

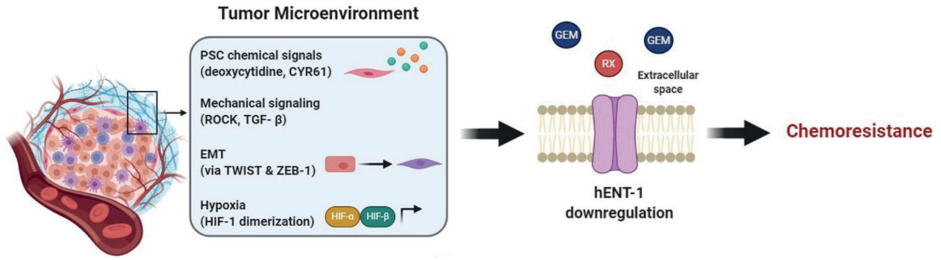


Figure 4. Tumor microenvironment affects hENT-1 expression causing chemoresistance. Components of the TME trigger PDAC hENT-1 downregulation and decreased activity. In order: PSCs secrete chemical factors, such as CYR61 and deoxycytidine, which downregulate hENT-1 and compete with drug metabolizing enzymes, respectively; mechanical signaling, such as TGF-β and ROCK, and EMT, triggered by TWIST and ZEB-1, also contribute to hENT-1 decreased activity; hypoxia induces HIF-α/HIF-B heterodimerization, which activates transcription of hypoxia-related genes further decreasing hENT-1 expression. Together these factors cause chemoresistance to drugs which are taken up via hENT-1 (e.g., gemcitabine (GEM) and RX-3117).

6.3.2. Mechanobiology

PDAC stroma, which accounts for the majority of tumor mass, is involved in resistance to gemcitabine, and a recent publication by Dalin et al. suggested a role of PSCs through an indirect influence on hENT-1. In this case the stroma compartment does not directly affect hENT-1 expression, but most likely bypasses it by producing elevated amounts of deoxycytidine. High deoxycytidine levels are taken up by PDAC cells, via hENT-1, and compete with gemcitabine's intracellular pool for phosphorylation and activation of the drug, therefore causing resistance [92]. Although further investigations are necessary [93] these promising preliminary results suggest that hENT-1 activity is indirectly influenced by the tumor stroma and that this interaction has a relevant role for PDAC gemcitabine chemoresistance.

Mechanical cues and hENT-1 are also intertwined, playing an additional role in PDAC EMT and resistance to gemcitabine. EMT is characterized by cellular physical changes, e.g., viscoelastic and stiffness properties modulation [94]. Notably, hENT-1 has been reported as a regulator of cellular mechanical properties, by means of EMT induction. Indeed, knockdown of hENT-1 was shown to induce cell elongation, stiffness reduction, increased migration potential and expression of EMT markers [95]. Consistently, recent studies reported that gemcitabine resistant cells, which had hENT-1 downregulation, were accompanied by an EMT phenotype. Additionally, these studies highlighted that the EMT was triggered by TWIST and ZEB-1 transcription factors. By inhibiting TWIST and ZEB-1 the cells had an increased hENT-1 expression and reversed EMT phenotype, increasing gemcitabine efficacy [96,97]. These interesting results suggest that reversing cellular mechanical changes, i.e., EMT, could at least in part contrast the phenomenon of gemcitabine resistance relying on hENT-1.

Lastly, mechanical signaling, that is signaling induced by or causing mechanical stress is also involved in hENT-1 regulation. In lung cancer, hENT-1 together with ROCK1-Rho A—kinases involved in actin organization, cell contractility and motility—are regulated by miR-26b. miR-26b mimic indeed is responsible for downregulation of the aforementioned proteins, leading to reduced tumor invasion and migration [98]. The exact mechanism behind hENT-1-ROCK interaction and whether this is also valid in PDAC has yet to be confirmed. Nevertheless, since both hENT-1 and ROCK are generally overexpressed in PDAC it is not a surprise that they are involved in the aggressive behavior and chemoresistance of this tumor type. Moreover, experimental data showed that in PDAC at least one mechanical signaling pathway is responsible for hENT-1 regulation. PSCs in the tumor microenvironment are a source of cysteine-rich angiogenic inducer 61 (CYR61) and this is due to upregulated TGF-β signaling, a pathway involved in mechanical signaling, ECM remodeling

and motility. Hesler et al. reported that this high concentration of CYR61 affected PDAC cells by downregulating hENT-1, therefore causing gemcitabine chemoresistance [99].

In conclusion, emerging evidence highlight the interaction of hENT-1 with tumor stroma and mechanical signaling, yet more evidence has to be obtained to have potential new targets to overcome pancreatic cancer progression and chemoresistance (Figure 4).

7. Discussion

The increasing use of a non-gemcitabine-based therapeutic option (FOLFIRINOX), both in the palliative and adjuvant setting, should prompt further development of predictive biomarkers for gemcitabine-based regimens. The validation of these biomarkers may indeed pave the road to the selection of patients who are likely to receive a benefit from gemcitabine-based adjuvant chemotherapy, and therefore has an immediate clinical relevance.

Remarkably, results obtained in several studies demonstrate that hENT-1 expression level is predictive of gemcitabine benefit, but only when assessed with the 10D7G2 mouse clone. Nevertheless, since this clone is not commercially available, this approach is not clinically-feasible and other strategies for the evaluation of hENT-1 status should be investigated. The quantification of hENT-1 mRNA levels could represent an alternative technique to IHC for the clinical evaluation of hENT-1 status in PDAC patients. A predictive role for hENT-1 mRNA expression in the treatment of PDAC with gemcitabine was previously reported in PDAC laser-microdissected specimens [42]. However, Raffenne and collaborators found that the threshold required to achieve statistical significance was higher with qRT-PCR compared with IHC, probably because no laser-microdissection was performed, leading to heavy microenvironment contamination and falsely decreased mRNA level [3]. In this regard, future studies involving laser microdissection techniques and larger cohorts of patients are required to validate the analysis of mRNA as a potential strategy for the clinical assessment of hENT-1 status.

Several additional factors are involved in cancer cell sensitivity to gemcitabine. For instance, the expression level and activity of gemcitabine-activating enzymes such as deoxycytidine kinase (dCK), and inactivating enzymes such as cytidine deaminase (CDA) and nucleotidase (NT5C1A/NT5C3) may provide a rational explanation for the discrepancy between high hENT-1 level and the poor response to gemcitabine-based chemotherapies [46]. In this regard, the study of more complex gemcitabine sensitivity/resistant signatures using both preclinical models and clinical samples, as well as novel technologies is warranted. For instance, a recent study on proteomics in gemcitabine resistant PANC1 cells and xenografts did not show a significant increase ($p = 0.29$) of hENT-1 expression in the resistant cells, though the fold-change was 1.5, while revealing that proteins associated with gemcitabine resistance are correlated with microtubule regulation [100]. Notably, this data provides an explanation as to why the combination of gemcitabine with nab-paclitaxel is effective in PDAC patients. However, hENT-1 is trafficked to the plasma membrane in association with microtubules suggesting a potential correlation with this system [101]. These findings are extremely interesting because another main challenge for the use of hENT-1 as a clinical biomarker is its validation in patients undergoing treatments with combination of gemcitabine and nab-paclitaxel and not only with gemcitabine monotherapy.

The role played by different cancer cell subpopulations, such as CSCs, as well as non-cancerous cells within the TME, in gemcitabine efficacy represents another critical issue, requiring further elucidation. Of note, recent studies showed that extracellular vesicles (EVs) can confer resistance to gemcitabine by miR-155-mediated suppression of dCK, which catalyzes the rate-limiting reaction in gemcitabine activation [102]. Future studies should investigate whether miRNAs affecting hENT-1 expression could also be involved in the transfer of a resistant phenotype through EVs.

The reliability of a predictive biomarker is assessed through the analysis of sensitivity, specificity, and positive and negative predictive values. However, these parameters need extensive validation studies and quality assessment prior to approval and application in clinical setting. Therefore, the expression levels of hENT-1 should be evaluated within trials testing therapeutic strategies/drugs than can bypass hENT-1 mediated gemcitabine resistance. For instance, NUC-1031 (Acelarin®, NuCana,

Edinburgh, UK), which is the first anti-cancer ProTide to enter the clinic, is a phosphoramidate modification of gemcitabine designed to overcome several mechanisms affecting gemcitabine efficacy. According to pre-clinical data, the increased hydrophobicity of NUC-1031 allows it to enter the cells bypassing the hENT-1 transporter [6], similarly to the previously tested lipid-drug conjugate CO-101 (CP-4126) [103,104]. However, the trials on CO-101 resulted in a disappointing failure in PDAC patients. In particular, in a randomized prospective study in patients with untreated metastatic PDAC, CO-101 was not superior to gemcitabine in the low tumor hENT-1 (assessed with the SP120 antibody) and an adverse effect profile similar to gemcitabine [5] was found, such as recently described for NUC-1031. Of note, hENT-1 is not the only potential mechanism mediating the favorable effect of NUC-1031, since this drug does not require the phosphorylation to difluorodeoxycytidine monophosphate (dFdCMP) by dCK and it preserves higher concentrations of the active triphosphate metabolite (dFdCTP) than gemcitabine at equimolar doses inside the tumor cells [6,105]. However, in the future trials on this drug in PDAC patients, standardized IHC technique for the detection of hENT-1 would be essential to overcome the criticisms about the previous CO-101 trial.

Several studies support the association between no or low hENT-1 expression in tumors and poor response to gemcitabine in other cancers, including bladder, biliary tract, and lung cancers. Thus, the validation of hENT-1 standardized IHC techniques would be useful also for other cancer types. For instance, Matsumura and colleagues evaluated the predictive potential of hENT-1 expression in patients with metastatic bladder cancer treated with gemcitabine-cisplatin-based combination chemotherapy. The IHC results showed that hENT-1 was localized in the cytoplasm of bladder tumor cells, and patients with high hENT-1 expression levels had a significantly longer median survival (17.3 months) compared to patients with lower levels (11.6 months) [106]. Similar results were observed by the IHC analysis of hENT-1 in a panel of patients with advanced Biliary Tract Cancer (BTC). Moreover, this study suggested that hENT-1 mediates the intracellular transport not only of gemcitabine but also of capecitabine, because a subpopulation of BTC patients treated with these two drugs showed a correlation between hENT-1 and OS [107]. Finally, Oguri and colleagues evaluated the hENT-1 expression in non-small cell lung cancer (NSCLC) patients who received gemcitabine-containing chemotherapy, showing that the absence of hENT-1 expression may be useful to predict resistance to gemcitabine-containing chemotherapy in NSCLC [108]. Of note this study showed that the protein expression of hENT-1 in a panel of cell lines with acquired resistance to different drugs, including gemcitabine, cisplatin and paclitaxel, was similar to the expression levels in their respective parental cells. These results suggest that hENT-1 is important in inherent resistance, but does not have a role in acquired resistance. Moreover, this transporter might still be used as potential predictive biomarker of gemcitabine efficacy also after pretreatment with other drugs, such as after neoadjuvant regimens, which are gaining momentum in the multidisciplinary treatment of even potentially resectable PDAC [109].

8. Conclusions

In the present article, we explored the contradictory data related to the predictive value of hENT-1 for gemcitabine activity in PDAC. We also considered the issues related to commercial and not commercial antibodies for IHC and laser-microdissected specimens for PCR analysis. Finally, we discussed the potential impact of different biological mechanisms on hENT-1 regulation, supporting the need of integrating additional (tissue and circulating) biomarkers and further exploring the uncertainty regarding the clinical significance within larger prospective trials using standardized methodologies.

The emergence of "omics" technologies (i.e., genomics, transcriptomics, proteomics, and metabolomics) has encouraged the discovery of new biomarkers. However, the identification of solid and reproducible molecular markers is amongst the biggest challenges in personalized cancer medicine. Therefore, in the present article, we have also reported molecular mechanisms influencing

the expression and activity of hENT-1, because we reckon that the integration of existing molecular knowledge should help to adjust for clinical data heterogeneity and limitation.

Last but not least, as reported in the tale of Ali Baba, in order to discover the secret of the cave, he was at right place at right time, suggesting that a relentless pursuit of the goals will lead to achieving success. Thus, persistent and appropriate studies are needed in order to validate effective biomarkers and will hopefully guide the selection of the best (sequence of) anticancer therapy in PDAC patients.

Supplementary Materials: The following are available online at http://www.mdpi.com/2072-6694/12/11/3206/s1, Table S1: Clinicopathological characteristics of PDAC patients evaluated for hENT-1 mRNA levels. Table S2: List of MicroRNA Targeting hENT-1. Figure S1: Chemical structures of the drugs transported by nucleoside transporters. Figure S2: Studies evaluating hENT-1 expression levels in pancreatic tumors and normal specimens. Figure S3: Stromal and tumor makers for accurate dissection.

Author Contributions: F.P. and O.R. were the principal investigators and take primary responsibility for the paper; G.M., A.G. and E.G. wrote additional paragraphs and prepared the table and figures, B.P., S.C., D.S.K.L. and D.C. provided essential material and participated in the research design; G.M., F.P., I.G. and A.E.F. checked the tables and figures the paper; G.J.P., A.E.F., I.G. and E.G. edited the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by the following grants: CCA Foundation 2012, 2015 and 2018 grants (G.J.P., E.G.), KWF Dutch Cancer Society grants (KWF project#19571, E.G.) and AIRC/Start-Up grant (E.G.); Action Against Cancer, UK and No Surrender Cancer Trust (in memory of Jason Boas), UK (A.E.F.).

Acknowledgments: The Authors would like to thank Sofia Koustoulidou for her initial literature work on this topic, and Tessa Le Large for the images of laser-microdissected PDAC specimens.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Molina-Arcas, M.; Trigueros-Motos, L.; Casado, F.J.; Pastor-Anglada, M. Physiological and Pharmacological Roles of Nucleoside Transporter Proteins. *Nucleosides Nucleotides Nucleic Acids* **2008**, *27*, 769–778. [CrossRef]
2. Robey, R.W.; Pluchino, K.M.; Hall, M.D.; Fojo, A.T.; Bates, S.E.; Gottesman, M.M. Revisiting the role of ABC transporters in multidrug-resistant cancer. *Nat. Rev. Cancer* **2018**, *18*, 452–464. [CrossRef]
3. Raffenne, J.; Nicolle, R.; Puleo, F.; Le Corre, D.; Boyez, C.; Marechal, R.; Emile, J.F.; Demetter, P.; Bardier, A.; Laurent-Puig, P.; et al. hENT1 Testing in Pancreatic Ductal Adenocarcinoma: Are We Ready? A Multimodal Evaluation of hENT1 Status. *Cancers* **2019**, *11*, 1808. [CrossRef]
4. El Hassouni, B.; Li Petri, G.; Liu, D.S.K.; Cascioferro, S.; Parrino, B.; Hassan, W.; Diana, P.; Ali, A.; Frampton, A.E.; Giovannetti, E. Pharmacogenetics of treatments for pancreatic cancer. *Expert Opin. Drug Metab. Toxicol.* **2019**, *15*, 437–447. [CrossRef]
5. Poplin, E.; Wasan, H.; Rolfe, L.; Raponi, M.; Ikdahl, T.; Bondarenko, I.; Davidenko, I.; Bondar, V.; Garin, A.; Boeck, S.; et al. Randomized, Multicenter, Phase II Study of CO-101 Versus Gemcitabine in Patients With Metastatic Pancreatic Ductal Adenocarcinoma: Including a Prospective Evaluation of the Role of hENT1 in Gemcitabine or CO-101 Sensitivity. *J. Clin. Oncol.* **2013**, *31*, 4453–4461. [CrossRef]
6. Blagden, S.P.; Rizzuto, I.; Suppiah, P.; O'Shea, D.; Patel, M.; Spiers, L.; Sukumaran, A.; Bharwani, N.; Rockall, A.; Gabra, H.; et al. Anti-tumour activity of a first-in-class agent NUC-1031 in patients with advanced cancer: Results of a phase I study. *Br. J. Cancer* **2018**, *119*, 815–822. [CrossRef]
7. Grossberg, A.J.; Chu, L.C.; Deig, C.R.; Fishman, E.K.; Hwang, W.L.; Maitra, A.; Marks, D.L.; Mehta, A.; Nabavizadeh, N.; Simeone, D.M.; et al. Multidisciplinary standards of care and recent progress in pancreatic ductal adenocarcinoma. *CA Cancer J. Clin.* **2020**. [CrossRef]
8. Collisson, E.A.; Bailey, P.; Chang, D.K.; Biankin, A.V. Molecular subtypes of pancreatic cancer. *Nat. Rev. Gastroenterol. Hepatol.* **2019**, *16*, 207–220. [CrossRef]
9. Kleeff, J.; Korc, M.; Apte, M.; La Vecchia, C.; Johnson, C.D.; Biankin, A.V.; Neale, R.E.; Tempero, M.; Tuveson, D.A.; Hruban, R.H.; et al. Pancreatic cancer. *Nat. Rev. Dis. Primers* **2016**, *2*, 16022. [CrossRef]
10. Giovannetti, E.; van der Borden, C.L.; Frampton, A.E.; Ali, A.; Firuzi, O.; Peters, G.J. Never let it go: Stopping key mechanisms underlying metastasis to fight pancreatic cancer. *Semin. Cancer Biol.* **2017**, *44*, 43–59. [CrossRef]

11. Adamska, A.; Domenichini, A.; Falasca, M. Pancreatic Ductal Adenocarcinoma: Current and Evolving Therapies. *Int. J. Mol. Sci.* **2017**, *18*, 1338. [CrossRef]
12. Neoptolemos, J.P.; Kleeff, J.; Michl, P.; Costello, E.; Greenhalf, W.; Palmer, D.H. Therapeutic developments in pancreatic cancer: Current and future perspectives. *Nat. Rev. Gastroenterol. Hepatol.* **2018**, *15*, 333–348. [CrossRef]
13. Mizrahi, J.D.; Surana, R.; Valle, J.W.; Shroff, R.T. Pancreatic cancer. *Lancet* **2020**, *395*, 2008–2020. [CrossRef]
14. Caparello, C.; Meijer, L.L.; Garajova, I.; Falcone, A.; Le Large, T.Y.; Funel, N.; Kazemier, G.; Peters, G.J.; Vasile, E.; Giovannetti, E. FOLFIRINOX and translational studies: Towards personalized therapy in pancreatic cancer. *World J. Gastroenterol.* **2016**, *22*, 6987–7005. [CrossRef] [PubMed]
15. Zheng, X.; Carstens, J.L.; Kim, J.; Scheible, M.; Kaye, J.; Sugimoto, H.; Wu, C.-C.; LeBleu, V.S.; Kalluri, R. Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. *Nature* **2015**, *527*, 525–530. [CrossRef] [PubMed]
16. Valle, S.; Martin-Hijano, L.; Alcalá, S.; Alonso-Nocelo, M.; Sainz, B., Jr. The Ever-Evolving Concept of the Cancer Stem Cell in Pancreatic Cancer. *Cancers* **2018**, *10*, 33. [CrossRef]
17. Zeng, S.; Pöttler, M.; Lan, B.; Grützmann, R.; Pilarsky, C.; Yang, H. Chemoresistance in Pancreatic Cancer. *Int. J. Mol. Sci.* **2019**, *20*, 4504. [CrossRef] [PubMed]
18. Principe, D.R.; DeCant, B.; Mascarinas, E.; Wayne, E.A.; Diaz, A.M.; Akagi, N.; Hwang, R.; Pasche, B.; Dawson, D.W.; Fang, D.; et al. TGF Signaling in the Pancreatic Tumor Microenvironment Promotes Fibrosis and Immune Evasion to Facilitate Tumorigenesis. *Cancer Res.* **2016**, *76*, 2525–2539. [CrossRef] [PubMed]
19. Pastor-Anglada, M.; Pérez-Torras, S. Emerging Roles of Nucleoside Transporters. *Front. Pharmacol.* **2018**, *9*, 606. [CrossRef]
20. Young, J.D.; Yao, S.Y.M.; Sun, L.; Cass, C.E.; Baldwin, S.A. Human equilibrative nucleoside transporter (ENT) family of nucleoside and nucleobase transporter proteins. *Xenobiotica* **2008**, *38*, 995–1021. [CrossRef]
21. King, A.E.; Ackley, M.A.; Cass, C.E.; Young, J.D.; Baldwin, S.A. Nucleoside transporters: From scavengers to novel therapeutic targets. *Trends Pharmacol. Sci.* **2006**, *27*, 416–425. [CrossRef]
22. Govindarajan, R.; Bakken, A.H.; Hudkins, K.L.; Lai, Y.; Casado, F.J.; Pastor-Anglada, M.; Tse, C.-M.; Hayashi, J.; Unadkat, J.D. In situ hybridization and immunolocalization of concentrative and equilibrative nucleoside transporters in the human intestine, liver, kidneys, and placenta. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **2007**, *293*, R1809–R1822. [CrossRef]
23. Pastor-Anglada, M.; Urtasun, N.; Perez-Torras, S. Intestinal Nucleoside Transporters: Function, Expression, and Regulation. *Compr. Physiol.* **2018**, *8*, 1003–1017. [CrossRef]
24. Rauchwerger, D.R.; Firby, P.S.; Hedley, D.W.; Moore, M.J. Equilibrative-sensitive nucleoside transporter and its role in gemcitabine sensitivity. *Cancer Res.* **2000**, *60*, 6075–6079.
25. Zhou, M.; Duan, H.; Engel, K.; Wang, J. Adenosine transport by plasma membrane monoamine transporter: Reinvestigation and comparison with organic cations. *Drug Metab. Dispos.* **2010**, *38*, 1798–1805. [CrossRef]
26. Boswell-Casteel, R.C.; Hays, F.A. Equilibrative nucleoside transporters—A review. *Nucleosides Nucleotides Nucleic Acids* **2017**, *36*, 7–30. [CrossRef]
27. Damaraju, V.L.; Damaraju, S.; Young, J.D.; Baldwin, S.A.; Mackey, J.; Sawyer, M.B.; Cass, C.E. Nucleoside anticancer drugs: The role of nucleoside transporters in resistance to cancer chemotherapy. *Oncogene* **2003**, *22*, 7524–7536. [CrossRef]
28. Balboni, B.; El Hassouni, B.; Honeywell, R.J.; Sarkisjan, D.; Giovannetti, E.; Poore, J.; Heaton, C.; Peterson, C.; Benaim, E.; Lee, Y.B.; et al. RX-3117 (fluorocyclopentenyl cytosine): A novel specific antimetabolite for selective cancer treatment. *Expert Opin. Investig. Drugs.* **2019**, *28*, 311–322. [CrossRef]
29. Wei, R.; Gust, S.L.; Tandio, D.; Maheux, A.; Nguyen, K.H.; Wang, J.; Bourque, S.; Plane, F.; Hammond, J.R. Deletion of murine *slc29a4* modifies vascular responses to adenosine and 5-hydroxytryptamine in a sexually dimorphic manner. *Physiol. Rep.* **2020**, *8*. [CrossRef]
30. Wright, N.J.; Lee, S.-Y. Structures of human ENT1 in complex with adenosine reuptake inhibitors. *Nat. Struct. Mol. Biol.* **2019**, *26*, 599–606. [CrossRef]
31. Elnaggar, M.; Giovannetti, E.; Peters, G.J. Molecular Targets of Gemcitabine Action: Rationale for Development of Novel Drugs and Drug Combinations. *Curr. Pharm. Des.* **2012**, *18*, 2811–2829. [CrossRef] [PubMed]
32. Cass, C.E.; Young, J.D.; Baldwin, S.A.; Cabrita, M.A.; Graham, K.A.; Griffiths, M.; Jennings, L.L.; Mackey, J.R.; Ng, A.M.L.; Ritzel, M.W.L.; et al. Nucleoside transporters of mammalian cells. *Pharm. Biotechnol.* **1999**, *12*, 313–352. [CrossRef]

33. GEPIA. Available online: http://gepia.cancer-pku.cn/detail.php?gene=SLC29A1 (accessed on 28 October 2020).
34. Griffiths, M.; Beaumont, N.; Yao, S.Y.M.; Sundaram, M.; Boumah, C.E.; Davies, A.; Kwong, F.Y.P.; Coe, I.; Cass, C.E.; Young, J.D.; et al. Cloning of a human nucleoside transporter implicated in the Cellular uptake of adenosine and chemotherapeutic drugs. *Nat. Med.* **1997**, *3*, 89–93. [CrossRef]
35. Mackey, J.R.; Mani, R.S.; Selner, M.; Mowles, D.; Young, J.D.; Belt, J.A.; Crawford, C.R.; Cass, C.E. Functional nucleoside transporters are required for gemcitabine influx and manifestation of toxicity in cancer cell lines. *Cancer Res.* **1998**, *58*, 4349–4357. [PubMed]
36. de Sousa Cavalcante, L.; Monteiro, G. Gemcitabine: Metabolism and molecular mechanisms of action, sensitivity and chemoresistance in pancreatic cancer. *Eur. J. Pharmacol.* **2014**, *741*, 8–16. [CrossRef]
37. Achiwa, H.; Oguri, T.; Sato, S.; Maeda, H.; Niimi, T.; Ueda, R. Determinants of sensitivity and resistance to gemcitabine: The roles of human equilibrative nucleoside transporter 1 and deoxycytidine kinase in non-small cell lung cancer. *Cancer Sci.* **2004**, *95*, 753–757. [CrossRef]
38. Mey, V.; Giovannetti, E.; Braud, F.D.; Nannizzi, S.; Curigliano, G.; Verweij, F.; Cobelli, O.D.; Pece, S.; Tacca, M.D.; Danesi, R. In vitro synergistic cytotoxicity of gemcitabine and pemetrexed and pharmacogenetic evaluation of response to gemcitabine in bladder cancer patients. *Br. J Cancer* **2006**, *95*, 289–297. [CrossRef]
39. Damaraju, V.L.; Scriver, T.; Mowles, D.; Kuzma, M.; Ryan, A.J.; Cass, C.E.; Sawyer, M.B. Erlotinib, Gefitinib, and Vandetanib Inhibit Human Nucleoside Transporters and Protect Cancer Cells from Gemcitabine Cytotoxicity. *Clin. Cancer Res.* **2014**, *20*, 176–186. [CrossRef]
40. Hubeek, I.; Giovannetti, E.; Broekhuizen, A.J.F.; Pastor-Anglada, M.; Kaspers, G.J.L.; Peters, G.J. Immunocytochemical Detection of hENT1 and hCNT1 in normal tissues, lung cancer cell lines, and NSCLC patient samples. *Nucleosides Nucleotides Nucleic Acids* **2008**, *27*, 787–793. [CrossRef] [PubMed]
41. Spratlin, J.; Sangha, R.; Glubrecht, D.; Dabbagh, L.; Young, J.D.; Dumontet, C.; Cass, C.; Lai, R.; Mackey, J.R. The Absence of Human Equilibrative Nucleoside Transporter 1 Is Associated with Reduced Survival in Patients With Gemcitabine-Treated Pancreas Adenocarcinoma. *Clin. Cancer Res.* **2004**, *10*, 6956–6961. [CrossRef] [PubMed]
42. Giovannetti, E.; Del Tacca, M.; Mey, V.; Funel, N.; Nannizzi, S.; Ricci, S.; Orlandini, C.; Boggi, U.; Campani, D.; Del Chiaro, M.; et al. Transcription Analysis of Human Equilibrative Nucleoside Transporter-1 Predicts Survival in Pancreas Cancer Patients Treated with Gemcitabine. *Cancer Res.* **2006**, *66*, 3928–3935. [CrossRef]
43. Farrell, J.J.; Elsaleh, H.; Garcia, M.; Lai, R.; Ammar, A.; Regine, W.F.; Abrams, R.; Benson, A.B.; Macdonald, J.; Cass, C.E.; et al. Human Equilibrative Nucleoside Transporter 1 Levels Predict Response to Gemcitabine in Patients With Pancreatic Cancer. *Gastroenterology* **2009**, *136*, 187–195. [CrossRef]
44. Fujita, H.; Ohuchida, K.; Mizumoto, K.; Itaba, S.; Ito, T.; Nakata, K.; Yu, J.; Kayashima, T.; Souzaki, R.; Tajiri, T.; et al. Gene Expression Levels as Predictive Markers of Outcome in Pancreatic Cancer after Gemcitabine-Based Adjuvant Chemotherapy. *Neoplasia* **2010**, *12*, 807–817. [CrossRef] [PubMed]
45. Morinaga, S.; Nakamura, Y.; Watanabe, T.; Mikayama, H.; Tamagawa, H.; Yamamoto, N.; Shiozawa, M.; Akaike, M.; Ohkawa, S.; Kameda, Y.; et al. Immunohistochemical Analysis of Human Equilibrative Nucleoside Transporter-1 (hENT1) Predicts Survival in Resected Pancreatic Cancer Patients Treated with Adjuvant Gemcitabine Monotherapy. *Ann. Surg. Oncol.* **2012**, *19*, 558–564. [CrossRef] [PubMed]
46. Maréchal, R.; Bachet, J.; Mackey, J.R.; Dalban, C.; Demetter, P.; Graham, K.; Couvelard, A.; Svrcek, M.; Bardier–Dupas, A.; Hammel, P.; et al. Levels of Gemcitabine Transport and Metabolism Proteins Predict Survival Times of Patients Treated with Gemcitabine for Pancreatic Adenocarcinoma. *Gastroenterology* **2012**, *143*, 664–674. [CrossRef] [PubMed]
47. Murata, Y.; Hamada, T.; Kishiwada, M.; Ohsawa, I.; Mizuno, S.; Usui, M.; Sakurai, H.; Tabata, M.; Ii, N.; Inoue, H.; et al. Human equilibrative nucleoside transporter 1 expression is a strong independent prognostic factor in UICC T3-T4 pancreatic cancer patients treated with preoperative gemcitabine-based chemoradiotherapy. *J. Hepatobiliary Pancreat. Sci.* **2012**, *19*, 413–425. [CrossRef] [PubMed]
48. Nakagawa, N.; Murakami, Y.; Uemura, K.; Sudo, T.; Hashimoto, Y.; Kondo, N.; Sueda, T. Combined analysis of intratumoral human equilibrative nucleoside transporter 1 (hENT1) and ribonucleotide reductase regulatory subunit M1 (RRM1) expression is a powerful predictor of survival in patients with pancreatic carcinoma treated with adjuvant gemcitabine-based chemotherapy after operative resection. *Surgery* **2013**, *153*, 565–575. [CrossRef]
49. Nordh, S.; Ansari, D.; Andersson, R. hENT1 expression is predictive of gemcitabine outcome in pancreatic cancer: A systematic review. *World J. Gastroenterol.* **2014**, *20*, 8482–8490. [CrossRef]

50. Greenhalf, W.; Ghaneh, P.; Neoptolemos, J.P.; Palmer, D.H.; Cox, T.F.; Lamb, R.F.; Garner, E.; Campbell, F.; Mackey, J.R.; Costello, E.; et al. Pancreatic Cancer hENT1 expression and survival from gemcitabine in patients from the ESPAC-3 trial. *J. Natl. Cancer Inst.* **2014**, *106*. [CrossRef]
51. Bird, N.T.E.; Elmasry, M.; Jones, R.; Psarelli, E.; Dodd, J.; Malik, H.; Greenhalf, W.; Kitteringham, N.; Ghaneh, P.; Neoptolemos, J.P.; et al. Immunohistochemical hENT1 expression as a prognostic biomarker in patients with resected pancreatic ductal adenocarcinoma undergoing adjuvant gemcitabine-based chemotherapy. *Br. J. Surg.* **2017**, *104*, 328–336. [CrossRef]
52. Yamada, R.; Mizuno, S.; Uchida, K.; Yoneda, M.; Kanayama, K.; Inoue, H.; Murata, Y.; Kuriyama, N.; Kishiwada, M.; Usui, M.; et al. Human Equilibrative Nucleoside Transporter 1 Expression in Endoscopic Ultrasonography-Guided Fine-Needle Aspiration Biopsy Samples Is a Strong Predictor of Clinical Response and Survival in the Patients With Pancreatic Ductal Adenocarcinoma Undergoing Gemcitabine-Based Chemoradiotherapy. *Pancreas* **2016**, *45*, 761–771. [CrossRef]
53. Brandi, G.; Deserti, M.; Vasuri, F.; Farioli, A.; Degiovanni, A.; Palloni, A.; Frega, G.; Barbera, M.A.; Lorenzo, S.; Garajova, I.; et al. Membrane Localization of Human Equilibrative Nucleoside Transporter 1 in Tumor Cells May Predict Response to Adjuvant Gemcitabine in Resected Cholangiocarcinoma Patients. *Oncologist* **2016**, *21*, 600–607. [CrossRef] [PubMed]
54. Meijer, L.L.; Puik, J.R.; Peters, G.J.; Kazemier, G.; Giovannetti, E. hENT-1 Expression and Localization Predict Outcome After Adjuvant Gemcitabine in Resected Cholangiocarcinoma Patients. *Oncologist* **2016**, *21*. [CrossRef] [PubMed]
55. Tavolari, S.; Deserti, M.; Vasuri, F.; Curti, S.; Palloni, A.; Pinna, A.D.; Cescon, M.; Frega, G.; De Lorenzo, S.; Barbera, M.A.; et al. Membrane human equilibrative nucleoside transporter 1 is associated with a high proliferation rate and worse survival in resected intrahepatic cholangiocarcinoma patients not receiving adjuvant treatments. *Eur. J. Cancer.* **2019**, *106*, 160–170. [CrossRef]
56. Kawada, N.; Uehara, H.; Katayama, K.; Nakamura, S.; Takahashi, H.; Ohigashi, H.; Ishikawa, O.; Nagata, S.; Tomita, Y. Human equilibrative nucleoside transporter 1 level does not predict prognosis in pancreatic cancer patients treated with neoadjuvant chemoradiation including gemcitabine. *J. Hepatobiliary Pancreat. Sci.* **2012**, *19*, 717–722. [CrossRef]
57. Sinn, M.; Riess, H.; Sinn, B.V.; Stieler, J.M.; Pelzer, U.; Striefler, J.K.; Oettle, H.; Bahra, M.; Denkert, C.; Bläker, H.; et al. Human equilibrative nucleoside transporter 1 expression analysed by the clone SP 120 rabbit antibody is not predictive in patients with pancreatic cancer treated with adjuvant gemcitabine—Results from the CONKO-001 trial. *Eur. J. Cancer* **2015**, *51*, 1546–1554. [CrossRef] [PubMed]
58. Ormanns, S.; Heinemann, V.; Raponi, M.; Isaacson, J.; Laubender, R.P.; Haas, M.; Kruger, S.; Kleespies, A.; Mann, E.; Bartosiewicz, M.; et al. Human equilibrative nucleoside transporter 1 is not predictive for gemcitabine efficacy in advanced pancreatic cancer: Translational results from the AIO-PK0104 phase III study with the clone SP120 rabbit antibody. *Eur. J. Cancer* **2014**, *50*, 1891–1899. [CrossRef]
59. Svrcek, M.; Cros, J.; Maréchal, R.; Bachet, J.-B.; Fléjou, J.-F.; Demetter, P. Human equilibrative nucleoside transporter 1 testing in pancreatic ductal adenocarcinoma: A comparison between murine and rabbit antibodies. *Histopathology* **2015**, *66*, 457–462. [CrossRef]
60. Kalloger, S.E.; Riazy, M.; Tessier-Cloutier, B.; Karasinska, J.M.; Gao, D.; Peixoto, R.D.; Samimi, S.; Chow, C.; Wong, H.-L.; Mackey, J.R.; et al. A predictive analysis of the SP120 and 10D7G2 antibodies for human equilibrative nucleoside transporter 1 (hENT1) in pancreatic ductal adenocarcinoma treated with adjuvant gemcitabine. *J. Pathol. Clin. Res.* **2017**, *3*, 179–190. [CrossRef]
61. Funel, N.; Giovannetti, E.; Pollina, L.E.; del Chiaro, M.; Mosca, F.; Boggi, U.; Campani, D. Critical role of laser microdissection for genetic, epigenetic and proteomic analyses in pancreatic cancer. *Expert Rev. Mol. Diagn.* **2011**, *11*, 695–701. [CrossRef]
62. Le Large, T.Y.S.; Mantini, G.; Meijer, L.L.; Pham, T.V.; Funel, N.; van Grieken, N.C.T.; Kok, B.; Knol, J.; van Laarhoven, H.W.M.; Piersma, S.R.; et al. Microdissected pancreatic cancer proteomes reveal tumor heterogeneity and therapeutic targets. *JCI Insight* **2020**, *5*, e138290. [CrossRef] [PubMed]
63. Jiraskova, L.; Ryska, A.; Duintjer Tebbens, E.J.; Hornychova, H.; Cecka, F.; Staud, F.; Cerveny, L. Are ENT1/ENT1, NOTCH3, and miR-21 Reliable Prognostic Biomarkers in Patients with Resected Pancreatic Adenocarcinoma Treated with Adjuvant Gemcitabine Monotherapy? *Cancers* **2019**, *11*, 1621. [CrossRef] [PubMed]

64. Mao, Y.; Shen, J.; Lu, Y.; Lin, K.; Wang, H.; Li, Y.; Chang, P.; Walker, M.G.; Li, D. RNA sequencing analyses reveal novel differentially expressed genes and pathways in pancreatic cancer. *Oncotarget* **2017**, *8*, 42537–42547. [CrossRef] [PubMed]
65. Zhang, G.; Schetter, A.; He, P.; Funamizu, N.; Gaedcke, J.; Ghadimi, B.M.; Ried, T.; Hassan, R.; Yfantis, H.G.; Lee, D.H.; et al. DPEP1 inhibits tumor cell invasiveness, enhances chemosensitivity and predicts clinical outcome in pancreatic ductal adenocarcinoma. *PLoS ONE* **2012**, *7*, e31507. [CrossRef] [PubMed]
66. Song, Y.; Wang, Q.; Wang, D.; Li, J.; Yang, J.; Li, H.; Wang, X.; Jin, X.; Jing, R.; Yang, J.-H.; et al. Label-Free Quantitative Proteomics Unravels Carboxypeptidases as the Novel Biomarker in Pancreatic Ductal Adenocarcinoma. *Transl. Oncol.* **2018**, *11*, 691–699. [CrossRef]
67. Mantini, G.; Vallés, A.M.; Le Large, T.Y.S.; Capula, M.; Funel, N.; Pham, T.V.; Piersma, S.R.; Kazemier, G.; Bijlsma, M.F.; Giovannetti, E.; et al. Co-expression analysis of pancreatic cancer proteome reveals biology and prognostic biomarkers. *Cell Oncol.* **2020**. [CrossRef]
68. Jaramillo, A.C.; Hubeek, I.; Broekhuizen, R.; Pastor-Anglada, M.; Kaspers, G.J.L.; Jansen, G.; Cloos, J.; Peters, G.J. Expression of the nucleoside transporters hENT1 (SLC29) and hCNT1 (SLC28) in pediatric acute myeloid leukemia. *Nucleosides Nucleotides Nucleic Acids* **2020**, 1–10. [CrossRef]
69. Sundaram, M.; Yao, S.Y.M.; Ng, A.M.L.; Griffiths, M.; Cass, C.E.; Baldwin, S.A.; Young, J.D. Chimeric Constructs between Human and Rat Equilibrative Nucleoside Transporters (hENT1 and rENT1) Reveal hENT1 Structural Domains Interacting with Coronary Vasoactive Drugs. *J. Biol. Chem.* **1998**, *273*, 21519–21525. [CrossRef]
70. SenGupta, D.J.; Lum, P.Y.; Lai, Y.; Shubochkina, E.; Bakken, A.H.; Schneider, G.; Unadkat, J.D. A single glycine mutation in the equilibrative nucleoside transporter gene, hENT1, alters nucleoside transport activity and sensitivity to nitrobenzylthioinosine. *Biochemistry* **2002**, *41*, 1512–1519. [CrossRef]
71. Leabman, M.K.; Huang, C.C.; DeYoung, J.; Carlson, E.J.; Taylor, T.R.; de la Cruz, M.; Johns, S.J.; Stryke, D.; Kawamoto, M.; Urban, T.J.; et al. Natural variation in human membrane transporter genes reveals evolutionary and functional constraints. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 5896–5901. [CrossRef]
72. Myers, S.N.; Goyal, R.K.; Roy, J.D.; Fairfull, L.D.; Wilson, J.W.; Ferrell, R.E. Functional single nucleotide polymorphism haplotypes in the human equilibrative nucleoside transporter 1. *Pharmacogenet. Genom.* **2006**, *16*, 315–320. [CrossRef] [PubMed]
73. Soo, R.A.; Wang, L.Z.; Ng, S.S.; Chong, P.Y.; Yong, W.P.; Lee, S.C.; Liu, J.J.; Choo, T.B.; Tham, L.S.; Lee, H.S.; et al. Distribution of gemcitabine pathway genotypes in ethnic Asians and their association with outcome in non-small cell lung cancer patients. *Lung Cancer* **2009**, *63*, 121–127. [CrossRef]
74. Gomez, A.; Ingelman-Sundberg, M. Pharmacoepigenetics: Its role in interindividual differences in drug response. *Clin. Pharmacol. Ther.* **2009**, *85*, 426–430. [CrossRef] [PubMed]
75. Manuyakorn, A.; Paulus, R.; Farrell, J.; Dawson, N.A.; Tze, S.; Cheung-Lau, G.; Hines, O.J.; Reber, H.; Seligson, D.B.; Horvath, S.; et al. Cellular histone modification patterns predict prognosis and treatment response in resectable pancreatic adenocarcinoma: Results from RTOG 9704. *J. Clin. Oncol.* **2010**, *28*, 1358–1365. [CrossRef] [PubMed]
76. Candelaria, M.; de la Cruz-Hernandez, E.; Taja-Chayeb, L.; Perez-Cardenas, E.; Trejo-Becerril, C.; Gonzalez-Fierro, A.; Chavez-Blanco, A.; Soto-Reyes, E.; Dominguez, G.; Trujillo, J.E.; et al. DNA Methylation-Independent Reversion of Gemcitabine Resistance by Hydralazine in Cervical Cancer Cells. *PLoS ONE* **2012**, *7*, e29181. [CrossRef]
77. Calin, G.A.; Croce, C.M. MicroRNA signatures in human cancers. *Nat. Rev. Cancer* **2006**, *6*, 857–866. [CrossRef]
78. Ji, Q.; Hao, X.; Zhang, M.; Tang, W.; Yang, M.; Li, L.; Xiang, D.; DeSano, J.T.; Bommer, G.T.; Fan, D.; et al. MicroRNA miR-34 inhibits human pancreatic cancer tumor-initiating cells. *PLoS ONE* **2009**, *4*, e6816. [CrossRef]
79. Lee, E.J.; Gusev, Y.; Jiang, J.; Nuovo, G.J.; Lerner, M.R.; Frankel, W.L.; Morgan, D.L.; Postier, R.G.; Brackett, D.J.; Schmittgen, T.D. Expression profiling identifies microRNA signature in pancreatic cancer. *Int. J. Cancer* **2006**, *120*, 1046–1054. [CrossRef]
80. Xu, Y.-F.; Hannafon, B.N.; Zhao, Y.D.; Postier, R.G.; Ding, W.-Q. Plasma exosome miR-196a and miR-1246 are potential indicators of localized pancreatic cancer. *Oncotarget* **2017**, *8*, 77028–77040. [CrossRef]
81. Chen, D.; Wu, X.; Xia, M.; Wu, F.; Ding, J.; Jiao, Y.; Zhan, Q.; An, F. Upregulated exosomic miR-23b-3p plays regulatory roles in the progression of pancreatic cancer. *Oncol. Rep.* **2017**, *38*, 2182–2188. [CrossRef]

82. Mikamori, M.; Yamada, D.; Eguchi, H.; Hasegawa, S.; Kishimoto, T.; Tomimaru, Y.; Asaoka, T.; Noda, T.; Wada, H.; Kawamoto, K.; et al. MicroRNA-155 Controls Exosome Synthesis and Promotes Gemcitabine Resistance in Pancreatic Ductal Adenocarcinoma. *Sci. Rep.* **2017**, *7*, 42339. [CrossRef] [PubMed]
83. Zhao, Y.; Zhao, L.; Ischenko, I.; Bao, Q.; Schwarz, B.; Nieß, H.; Wang, Y.; Renner, A.; Mysliwietz, J.; Jauch, K.-W.; et al. Antisense inhibition of microRNA-21 and microRNA-221 in tumor-initiating stem-like cells modulates tumorigenesis, metastasis, and chemotherapy resistance in pancreatic cancer. *Targeted Oncol.* **2015**, *10*, 535–548. [CrossRef] [PubMed]
84. Weniger, M.; Honselmann, K.; Liss, A. The Extracellular Matrix and Pancreatic Cancer: A Complex Relationship. *Cancers* **2018**, *10*, 316. [CrossRef]
85. Jain, R.K. Normalization of tumor vasculature: An emerging concept in antiangiogenic therapy. *Science* **2005**, *307*, 5758–5762. [CrossRef]
86. Koong, A.C.; Mehta, V.K.; Le, Q.T.; Fisher, G.A.; Terris, D.J.; Brown, J.M.; Bastidas, A.J.; Vierra, M. Pancreatic tumors show high levels of hypoxia. *Int. J. Radiat. Oncol. Biol. Phys.* **2000**, *48*, 919–922. [CrossRef]
87. Büchler, P.; Reber, H.A.; Büchler, M.; Shrinkante, S.; Büchler, M.W.; Friess, H.; Semenza, G.L.; Hines, O.J. Hypoxia-inducible factor 1 regulates vascular endothelial growth factor expression in human pancreatic cancer. *Pancreas* **2003**, *26*, 56–64. [CrossRef]
88. Shibaji, T.; Nagao, M.; Ikeda, N.; Kanehiro, H.; Hisanaga, M.; Ko, S.; Fukumoto, A.; Nakajima, Y. Prognostic significance of HIF-1 alpha overexpression in human pancreatic cancer. *Anticancer Res.* **2003**, *23*, 4721–4727.
89. Longo, V.; Brunetti, O.; Gnoni, A.; Cascinu, S.; Gasparini, G.; Lorusso, V.; Ribatti, D.; Silvestris, N. Angiogenesis in pancreatic ductal adenocarcinoma: A controversial issue. *Oncotarget* **2016**, *7*, 58649–58658. [CrossRef] [PubMed]
90. Eltzschig, H.K.; Abdulla, P.; Hoffman, E.; Hamilton, K.E.; Daniels, D.; Schönfeld, C.; Löffler, M.; Reyes, G.; Duszenko, M.; Karhausen, J.; et al. HIF-1–dependent repression of equilibrative nucleoside transporter (ENT) in hypoxia. *J. Exp. Med.* **2005**, *202*, 1493–1505. [CrossRef]
91. Löffler, M.; Morote-Garcia, J.C.; Eltzschig, S.A.; Coe, I.R.; Eltzschig, H.K. Physiological roles of vascular nucleoside transporters. *Arterioscler. Thromb. Vasc. Biol.* **2007**, *27*, 1004–1013. [CrossRef]
92. Dalin, S.; Sullivan, M.R.; Lau, A.N.; Grauman-Boss, B.; Mueller, H.S.; Kreidl, E.; Fenoglio, S.; Luengo, A.; Lees, J.A.; Vander Heiden, M.G.; et al. Deoxycytidine Release from Pancreatic Stellate Cells Promotes Gemcitabine Resistance. *Cancer Res.* **2019**, *79*, 5723–5733. [CrossRef] [PubMed]
93. Che, P.P.; Gregori, A.; Firuzi, O.; Dahele, M.; Sminia, P.; Peters, G.J.; Giovannetti, E. Pancreatic cancer resistance conferred by stellate cells: Looking for new preclinical models. *Exp. Hematol. Oncol.* **2020**, *9*, 18. [CrossRef]
94. Mendez, M.G.; Restle, D.; Janmey, P.A. Vimentin enhances cell elastic behavior and protects against compressive stress. *Biophys. J.* **2014**, *107*, 314–323. [CrossRef] [PubMed]
95. Lee, Y.; Koay, E.J.; Zhang, W.; Qin, L.; Kirui, D.K.; Hussain, F.; Shen, H.; Ferrari, M. Human equilibrative nucleoside transporter-1 knockdown tunes cellular mechanics through epithelial-mesenchymal transition in pancreatic cancer cells. *PLoS ONE* **2014**, *9*, e107973. [CrossRef] [PubMed]
96. Yan, T.; Li, H.-Y.; Wu, J.-S.; Niu, Q.; Duan, W.-H.; Han, Q.-Z.; Ji, W.-M.; Zhang, T.; Lv, W. Astaxanthin inhibits gemcitabine-resistant human pancreatic cancer progression through EMT inhibition and gemcitabine resensitization. *Oncol. Lett.* **2017**, *14*, 5400–5408. [CrossRef] [PubMed]
97. Chen, S.; Wang, Y.; Zhang, W.-L.; Dong, M.-S.; Zhang, J.-H. Sclareolide enhances gemcitabine-induced cell death through mediating the NICD and Gli1 pathways in gemcitabine-resistant human pancreatic cancer. *Mol. Med. Rep.* **2017**, *15*, 1461–1470. [CrossRef]
98. Gao, Y.; Yang, F. MIR-26b regulates invasion and migration of lung cancer cells through targeting hENTI depending on RhoA/ROCK-I pathway. *Zhong Nan Da Xue Xue Bao Yi Xue Ban* **2017**, *42*, 755–761. [CrossRef]
99. Hesler, R.A.; Huang, J.J.; Starr, M.D.; Treboschi, V.M.; Bernanke, A.G.; Nixon, A.B.; McCall, S.J.; White, R.R.; Blobe, G.C. TGF-β-induced stromal CYR61 promotes resistance to gemcitabine in pancreatic ductal adenocarcinoma through downregulation of the nucleoside transporters hENT1 and hCNT3. *Carcinogenesis* **2016**, *37*, 1041–1051. [CrossRef]
100. Le Large, T.Y.S.; El Hassouni, B.; Funel, N.; Kok, B.; Piersma, S.R.; Pham, T.V.; Olive, K.P.; Kazemier, G.; van Laarhoven, H.W.M.; Jimenez, C.R.; et al. Proteomic analysis of gemcitabine-resistant pancreatic cancer cells reveals that microtubule-associated protein 2 upregulation associates with taxane treatment. *Ther. Adv. Med. Oncol.* **2019**, *11*. [CrossRef]

101. Nivillac, N.M.I.; Bacani, J.; Coe, I.R. The life cycle of human equilibrative nucleoside transporter 1: From ER export to degradation. *Exp. Cell Res.* **2011**, *317*, 1567–1579. [CrossRef]
102. Patel, G.K.; Khan, M.A.; Bhardwaj, A.; Srivastava, S.K.; Zubair, H.; Patton, M.C.; Singh, S.; Khushman, M.; Singh, A.P. Exosomes confer chemoresistance to pancreatic cancer cells by promoting ROS detoxification and miR-155-mediated suppression of key gemcitabine-metabolising enzyme, DCK. *Br. J. Cancer* **2017**, *116*, 609–619. [CrossRef] [PubMed]
103. Bergman, A.M.; Adema, A.D.; Balzarini, J.; Bruheim, S.; Fichtner, I.; Noordhuis, P.; Fodstad, Ø.; Myhren, F.; Sandvold, M.L.; Hendriks, H.R.; et al. Antiproliferative activity, mechanism of action and oral antitumor activity of CP-4126, a fatty acid derivative of gemcitabine, in in vitro and in vivo tumor models. *Invest. New Drugs* **2011**, *29*, 456–466. [CrossRef] [PubMed]
104. Adema, A.D.; Smid, K.; Losekoot, N.; Honeywell, R.J.; Verheul, H.M.; Myhren, F.; Sandvold, M.L.; Peters, G.J. Metabolism and accumulation of the lipophilic deoxynucleoside analogs elacytarabine and CP-4126. *Investig. New Drugs* **2012**, *30*, 1908–1916. [CrossRef] [PubMed]
105. Rizzuto, I.; Ghazaly, E.; Peters, G.J. Pharmacological factors affecting accumulation of gemcitabine's active metabolite, gemcitabine triphosphate. *Pharmacogenomics* **2017**, *18*, 911–925. [CrossRef]
106. Matsumura, N.; Nakamura, Y.; Kohjimoto, Y.; Inagaki, T.; Nanpo, Y.; Yasuoka, H.; Ohashi, Y.; Hara, I. The prognostic significance of human equilibrative nucleoside transporter 1 expression in patients with metastatic bladder cancer treated with gemcitabine-cisplatin-based combination chemotherapy. *BJU Int.* **2011**, *108*, E110–E116. [CrossRef]
107. Santini, D.; Schiavon, G.; Vincenzi, B.E.; Cass, C.; Vasile, E.D.; Manazza, A.; Catalano, V.G.; Baldi, G.; Lai, R.; Rizzo, S.; et al. Human Equilibrative Nucleoside Transporter 1 (hENT1) Levels Predict Response to Gemcitabine in Patients With Biliary Tract Cancer (BTC). *Curr. Cancer Drug Targets* **2011**, *11*, 123–129. [CrossRef]
108. Oguri, T.; Achiwa, H.; Muramatsu, H.; Ozasa, H.; Sato, S.; Shimizu, S.; Yamazaki, H.; Eimoto, T.; Ueda, R. The absence of human equilibrative nucleoside transporter 1 expression predicts nonresponse to gemcitabine-containing chemotherapy in non-small cell lung cancer. *Cancer Lett.* **2007**, *256*, 112–119. [CrossRef]
109. Oba, A.; Ho, F.; Bao, Q.R.; Al-Musawi, M.H.; Schulick, R.D.; Del Chiaro, M. Neoadjuvant Treatment in Pancreatic Cancer. *Front. Oncol.* **2020**, *10*, 245. [CrossRef]

Supplementary Appendix Chapter 9

"Open Sesame?": Biomarker Status of the Human Equilibrative Nucleoside Transporter-1 and Molecular Mechanisms Influencing its Expression and Activity in the Uptake and Cytotoxicity of Gemcitabine in Pancreatic Cancer

Table S1. Clinicopathological Characteristics of PDAC Patients Evaluated for hENT-1 mRNA Levels. The expression of hENT-1 was evaluated by quantitative PCR, as described previously[1].

Age—years	
Mean (± SD)	65 (± 5)
Sex—No. (%)	
Male	12 (54.5)
Female	10 (45.5)
Stage *—No. (%)	
II	11 (50)
III	11 (50)
Grading (%)	
G1-G2§	9 (40.9)
G3	13 (59.1)

Notes: *AJCC Cancer Staging Manual, 7th Edition; WHO grading system 2007; Abbreviations: PDAC = pancreatic ductal adenocarcinoma, No. = number of patients.

Table S2. List of MicroRNA targeting hENT-1.

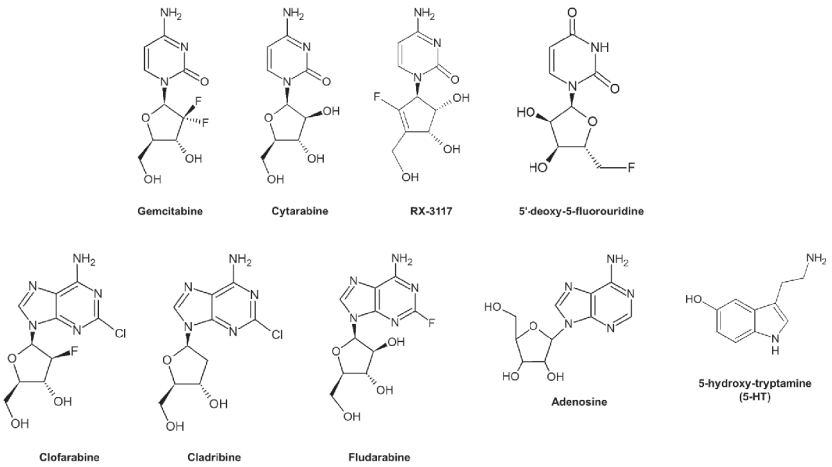
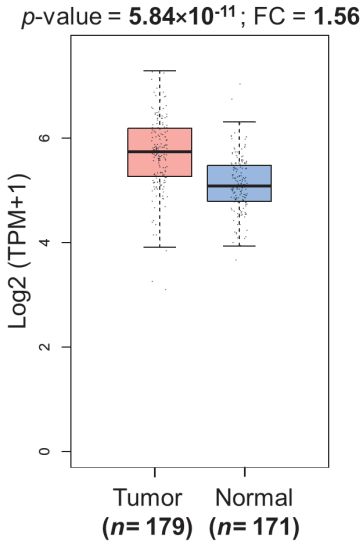


Figure S1. Chemical structures of the drugs transported by nucleoside transporters.

A



B

Studies on tumor and non-tumor matched samples reporting hENT-1 expression

Study	Analysis type	Database source	N samples	Reference
Mao *et al.* 2017	RNA-seq	paper suppl. info: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5522086/	10	[64]
Zhang G. *et al.* 2012	microarray	GSE28735	45	[65]
Song *et al.* 2018	nanoLC MS/MS	paper suppl. info: https://www.sciencedirect.com/science/article/pii/S1936523318300597	3	[66]

Figure S2. Studies evaluating hENT-1 expression levels in pancreatic tumors and normal specimens. The nucleoside transporter hENT-1iis overexpressed in different tumor types, including PDAC analyzing the RNA sequencing expression data of 9,736 tumors and 8,587 normal samples from the TCGA and the GTEx projects (http://gepia.cancer-pku.cn/detail.php?gene=SLC29A1) (panel A). However, the analysis of similarly matched transcriptomics and proteomics public datasets did not show a significance difference in hENT-1 expression levels (panel B).

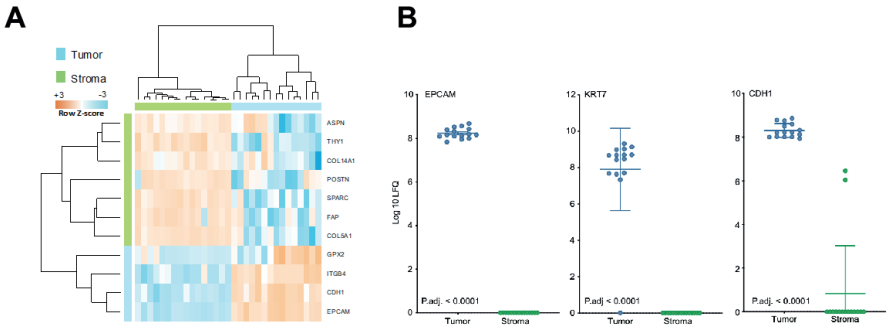


Figure S3. Stromal and tumor compartment dissection was confirmed by evaluating the protein expression of epithelial and stromal markers (panel **A**). Moreover, protein expression of specific tumor markers (EPCAM, KRT7 and CDH1) in matched stromal samples ranged from zero to very low levels, indicating a minimal stromal contamination (panel **B**).

Reference supplemental data

1. Toffalorio, F. *et al.* Expression of gemcitabine- and cisplatin-related genes in non-small-cell lung cancer. *Pharmacogenomics J.* **10**, 180–190 (2010).

Chapter 10

Interrelationship between miRNA and splicing factors in pancreatic ductal adenocarcinoma

Mantini G*, Sudupmadata D*, Randazzo O*, Capula M, Muller I.B., Cascioferro S, Diana P, Peters GJ, Giovannetti E.
*These authors contributed equally

Epigenetics, 2021

EPIGENETICS
https://doi.org/10.1080/15592294.2021.1916697

Taylor & Francis
Taylor & Francis Group

REVIEW

OPEN ACCESS Check for updates

Interrelationship between miRNA and splicing factors in pancreatic ductal adenocarcinoma

I Gede Putu Supadmanaba [a,b*], Giulia Mantini [a,c*], Ornella Randazzo [a,d*], Mjriam Capula[c,e], Ittai B. Muller[f], Stella Cascioferro[d], Patrizia Diana[d], Godefridus J. Peters [a,g], and Elisa Giovannetti [a,c]

[a]Department of Medical Oncology, Cancer Center Amsterdam, Amsterdam UMC, VU University Medical Center (VUMC), Amsterdam, The Netherlands; [b]Biochemistry Department, Faculty of Medicine, Universitas Udayana, Denpasar, Bali, Indonesia; [c]Cancer Pharmacology Lab, AIRC Start up Unit, Fondazione Pisana per La Scienza, Pisa, Italy; [d]Dipartimento Di Scienze E Tecnologie Biologiche Chimiche E Farmaceutiche (STEBICEF), Università Degli Studi Di Palermo, Palermo, Italy; [e]Institute of Life Sciences, Sant'Anna School of Advanced Studies, Pisa, Italy; [f]Department of Clinical Chemistry, Amsterdam UMC, VU University Medical Center (VUMC), Amsterdam, The Netherlands; [g]Department of Biochemistry, Medical University of Gdansk, Poland

ABSTRACT

Pancreatic ductal adenocarcinoma (PDAC) is one of the most lethal cancers because of diagnosis at late stage and inherent/acquired chemoresistance. Recent advances in genomic profiling and biology of this disease have not yet been translated to a relevant improvement in terms of disease management and patient's survival. However, new possibilities for treatment may emerge from studies on key epigenetic factors. Deregulation of microRNA (miRNA) dependent gene expression and mRNA splicing are epigenetic processes that modulate the protein repertoire at the transcriptional level. These processes affect all aspects of PDAC pathogenesis and have great potential to unravel new therapeutic targets and/or biomarkers. Remarkably, several studies showed that they actually interact with each other in influencing PDAC progression. Some splicing factors directly interact with specific miRNAs and either facilitate or inhibit their expression, such as Rbfox2, which cleaves the well-known oncogenic miRNA miR-21. Conversely, miR-15a-5p and miR-25-3p significantly downregulate the splicing factor hnRNPA1 which acts also as a tumour suppressor gene and is involved in processing of miR-18a, which in turn, is a negative regulator of KRAS expression. Therefore, this review describes the interaction between splicing and miRNA, as well as bioinformatic tools to explore the effect of splicing modulation towards miRNA profiles, in order to exploit this interplay for the development of innovative treatments. Targeting aberrant splicing and deregulated miRNA, alone or in combination, may hopefully provide novel therapeutic approaches to fight the complex biology and the common treatment recalcitrance of PDAC.

ARTICLE HISTORY
Received 5 December 2020
Revised 23 March 2021
Accepted 6 April 2021

KEYWORDS
PDAC; splicing deregulation; miRNA; interaction; splicing modulation

Introduction

Pancreatic ductal adenocarcinoma (PDAC) is one of the most lethal cancers worldwide [1,2]. Although its incidence and prevalence are lower than several other cancers, such as lung, head and neck, colorectal and breast cancer, the mortality rate almost matches the incidence rate [3]. Furthermore, early detection is difficult in PDAC because of the lack of accurate biomarkers [3,4]. Current biomarkers still have low sensitivity and specificity and, therefore, are not suitable as screening methods [5]. The problem becomes even more complicated as this type of cancer is hard to treat or to manage. PDAC has a rapid progression and only 20% of newly diagnosed patients are eligible for surgical resection, the most effective treatment option for this disease [1]. In addition, PDAC is highly resistant to any therapy upfront and tends also to rebound rapidly after first response/stabilization [3].

Recent studies provided new insights into the underlying mechanism of PDAC evolution, suggesting that, in addition to the specific mutational load, including the concurrent mutations in KRAS, TP53, p16, and DPC4, and to the tumour and stromal heterogeneity, microRNAs (miRNAs) and splicing deregulation could be major players in directing tumorigenesis and tumour evolution [6–9].

CONTACT Dr. Elisa Giovannetti e.giovannetti@amsterdamumc.nl Department of Medical Oncology, Amsterdam UMC VU_{mc}-location, De Boelelaan 1117, 1081 HV Amsterdam, The Netherlands

*equally contribution

MiRNAs have been studied extensively and there is a wide array of functionally relevant miRNAs that play an important role in PDACs [7]. For instance, miR-21 has been implicated in carcinogenesis and tumour progression in many types of solid cancers, including PDAC [10,11]. It may be a strong biomarker for early detection, but lacks specificity as it is upregulated in many types of solid cancers and other diseases [12]. Another well-known miRNA in PDAC is miR-155 that is important for inflammation and metastatic processes, while miR-121 and miR-21 contribute to chemoresistance [13].

Contrary to miRNA, splicing deregulation is a relatively new concept in cancer progression, especially in solid cancers [8]. The discovery of mutations in genes encoding splicing factors increased the interests in this topic, prompting several recent studies. Splicing deregulation due to mutation of splicing factors is especially important in non-solid cancers, such as leukaemia, but their overexpression is widely observed also in solid cancers [14,15]. Despite a low mutational rate of splicing factor 3B subunit 1 (SF3B1) in mesothelioma, its overexpression is diffusely prevalent and significantly associated with increased malignant characteristics and patients survival [16]. In lung cancer, Serine and Arginine Rich Splicing Factor 2 (SRSF2) has been implicated in patient's survival and tumour progression while Serine/arginine-rich splicing factor 7 (SRSF7) is highly expressed in chemoresistant colorectal cancer [17].

Although miRNA and splicing deregulation draw extensive interest among cancer scientists, the possibility of their interaction emerged only recently. Rodriguez-Aguayo et al. [18] showed that miR-15a-5p and miR-25-3p significantly downregulate tumour suppressor gene splicing factor hnRNPA1 which is a key player in the processing of miR-18a, the negative regulator of KRAS expression. On the other hand, splicing factors themselves could affect miRNA expression as shown by Chen et al., who reported a inhibition of miR-21 by Rbfox-2 [18]. These evidences indicate that there could be a relationship between miRNA and gene splicing, which might influence different oncogenic processes and provide new crucial concepts to be exploited towards more efficient cancer treatments. Therefore, this review discusses this intricate relationship, with a focus on the basic concept of miRNA and splicing deregulation, and on how miRNA and splicing factors affect each other. Additionally, this review concludes with consideration on how to exploit this relationship for future strategies in PDAC management and treatment.

Biology of miRNA and its relevance in PDAC

By definition, a miRNA is a non-coding RNA which typically consists of 19–24 nucleotides with a pivotal role in post-transcriptional regulation [7]. MiRNAs were first discovered in *Caenorhabditis elegans* and thousands more have been identified in all kinds of organisms [19]. In humans, around 2500 different miRNAs have been identified along with their sequence, transcript annotation, and their location within the genome [20]. MiRNAs biogenesis begins with their transcription by RNA polymerase II which generates long precursors known as primary miRNA transcripts (pri-miRNAs) (**Figure S1**). This transcript has a wide variability in their length, but it typically ranges between 100 and 1000 base pairs. The pri-miRNAs are then processed by Drosha-DGCR8 ribonuclease complex in the nucleus, producing 70–100 nucleotides long intermediate pre-miRNAs with hairpin shape. Then, this intermediate will be transported into the cytoplasm by Exportin-5 and RanGTP6 where it will further be processed by the endoribonuclease Dicer (also known as endoribonuclease RNase III). Dicer cleaves the terminal loop and produces mature double stranded 19–24 nts RNA. One strand of this mature miRNA will be degraded, while the other is incorporated into Argonaute heteromultimer protein to form highly specialized RNA Induced Silencing Complex (RISC) [21]. The seed sequence in miRNA leads to the complex towards target mRNA by means of RNA-RNA base pairing. The target mRNAs can either be degraded or translationally repressed, which depends on whether the seed sequence matches the target sequence within the mRNA. Complete base pairing between the seed sequences and target mRNAs usually leads to degradation of mRNA while incomplete pairing results in translational suppression [22]. However, this

phenomenon also underlies the reason why miRNA is so versatile. It has been shown that a single miRNA can indeed control mRNAs from several genes, while a single mRNA can be targeted by several different miRNAs [23]. In addition, it is estimated that 60% of all genes are controlled by miRNAs which further underscores their importance in the control of gene expression [24]. MiRNAs are also involved in important cellular processes, such as cell proliferation, metabolism, differentiation, apoptosis, and cell signalling[25].

Extensive studies on miRNA gave insight in its important role in many types of cancer, including PDAC. Cancer cells are known to have aberrant miRNA expression: where tumour suppressors' miRNAs are often downregulated, miRNAs promoting carcinogenesis or tumour progression is usually over-expressed [25]. This deregulation often leads to aberrant cellular processes, such as uncontrolled mitosis, apoptosis, drug resistance, invasion, metastasis and angiogenesis [7,13,14,25].

The first evidence of miRNA dysregulation in PDAC was reported by Poy et al. [26] through a profiling study using mouse pancreas. Follow-up studies using different types of samples confirmed the initial finding that PDAC has a specific miRNA expression profile [17,27,28]. In particular, a study comparing PDAC tissue with adjacent normal pancreatic tissue reported a total of 158 miRNAs differentially expressed [29]. Fifty-one miRNAs were upregulated including miR-196, miR-200a, miR-21, and miR-27a, while 107 miRNAs were downregulated, with miR-96, miR-200, and miR-217 being the most significant [29].

In Table 1 we report an overview of the clinical evidence on miRNA deregulation in PDAC as well as the most interesting preclinical findings on candidate miRNAs emerging from these studies. Schultz and colleagues [28] reported that 43 miRNAs were upregulated while 41 were downregulated when comparing paraffin-embedded PDAC tissue samples with normal ones. The expression of key miRNAs was also different between resectable and non-resectable PDAC patients as reported by Calatayud et al. [30]. Around 22 miRNAs were differentially expressed with miR-64, miR-136, miR-196, miR-492, and miR-622 being the most significant. A separate study by Papaconstantinou et al. showed a different but also some consistent results [31]; miR-21, miR-155, miR-205, miR-221, and miR-222 were consistently overexpressed while miR-31, miR-122, miR-146, and miR-375 were downregulated in PDAC samples. Preclinical and functional analysis showed that miR-21 and miR-155 are the only two miRNAs that were consistently overexpressed and linked to cancer progression [17]. However, several profiling studies showed consistent overexpression of miR-21, miR-155, and miR-221, while miR-34 and miR-145 were downregulated [32]. Remarkably, miR-21 and miR-155 obtained from pancreatic tissue could differentiate malignant from benign lesions with high accuracy [33] and have both been proven to be able to differentiate between pancreatic intra-epithelial neoplasia (PanIN) with normal pancreatic [34,35]. Clinically, miRNAs have been assessed to differentiate benign and malignant lesions, determining the stage of PDAC, as a biomarker for metastasis, and to predict the therapeutic outcome, as illustrated in Figure 1. The potential role of specific miRNA in early diagnostics is particularly important in PDAC since screening modalities are very limited and the disease tends to be diagnosed in advanced stage which has a high risk of metastasis and low therapeutic response [3]. Furthermore, miR-155 is increasingly expressed as early as PanIN-1 while miR-21 is beginning to be abundant in PanIN-2 and −3, suggesting that miR-21 is more suitable for advanced disease marker. Another microRNA that has increased expression in advanced PanIN (PanIN-3) is miR-196b while the expressions of miR-133, miR-185, miR-200 c, and miR-34 c are higher in low-grade neoplasia [36].

The expression levels of miR-21 and miR-155 were also up-regulated also in invasive intraductal papillary mucinous neoplasms (IPMNs) of the pancreas, compared to non-invasive IPMNs, as well as in non-invasive IPMNs compared with normal tissues. Conversely, miR-101 levels were significantly higher in non-invasive IPMNs and normal tissues compared with invasive IPMNs. Furthermore, miR-21 emerged as an independent prognostic biomarker in invasive IPMNs [37].

The information on circulating miRNAs in preneoplastic lesions is limited while circulating miR-10b, miR-20a, miR-21, miR-24, miR-25, miR-99a, miR-100b, miR-155, miR-185, and miR-191 showed a high accuracy in differentiating PDAC with pancreatitis and normal pancreas. Regarding metastasis, both miR-21 and miR-155 have been proven to actively play important role in inducing cell migration and

Table 1. MiRNAs aberrantly expressed in PDAC samples.

	Expression level (N)					
miRNA	Normal Tissue	PDAC	Tissue type	N Samples	Method	Reference
	Low	High	FFPE	Pancreatic cancer (n = 165); Normal Pancreas (n = 35)	RT-PCR	[30]
	Low	High	Fresh Tissue	Pancreatic Cancer (n = 88); Normal Pancreas (n = 98)	qRT-PCR	[31]
miR-21	Low	High	Biopsy	Metastatic (n = 31); Non-metastatic (n = 50)	q-PCR	[43]
	Low	High	FFPE	Adjuvant therapy (n = 52); Non-adjuvant therapy (n = 27)	qRT-PCR	[45]
	Low	High	FFPE	Normal pancreas (n = 12); Pancreatitis (n = 45); PDAC (n = 80)	In-situ Hybridization	[171]
	Low	High	Plasma	PDAC (n = 32); Normal healthy (n = 30)	qRT-PCR	[9]
	Low	High	Fresh Tissue	Pancreatic cancer (n = 88); Normal pancreas (n = 98)	qRT-PCR	[31]
miR-155	Low	High	FFPE	Pancreatic lesions (n = 55)	qRT-PCR	[33]
	Low	High	Plasma	Pancreatic cancer (n = 40); Normal pancreas (n = 25)	q-PCR	[172]
	Low (N = 80)	High (N = 80)	FFPE	Pancreatic cancer (n = 80); Normal pancreas (n = 80)	In-situ Hybridization	[173]
	Low (N = 98)	High (N = 88)	Fresh Tissue	Pancreatic cancer (n = 88); Normal pancreas (n = 98)	qRT-PCR	[30]
miR-205	Low (N = 17)	High (N = 34)	FFPE	Pancreatic cancer (n = 34); Normal pancreas (n = 17)	qRT-PCR	[173]
	Low (N = 17)	High (N = 47)	Serum	Pancreatic cancer (n = 47); Normal pancreas (n = 17)	qRT-PCR	[173]
miR-205	Low (N = 5)	High (N = 5)	Fresh Tissue	Pancreatic cancer (n = 5); Normal pancreas (n = 5)	MiRNA Microarray	[174]
miR-196b, miR-217, miR-411, miR-198	High (N = 28)	Low (N = 170)	FFPE	Pancreatic cancer (n = 170); Normal pancreas (n = 28)	RT-PCR	[28]
miR-210, miR-222	Low (N = 98)	High (N = 88)	Fresh Tissue	Pancreatic cancer (n = 88); Normal pancreas (n = 98)	qRT-PCR	[31]
miR-375	High (N = 35)	Low (N = 165)	FFPE	Pancreatic cancer (n = 165); Normal pancreas (n = 35)	RT-PCR	[30]
miR-377	High (N = 30)	Low (N = 30)	Snap-frozen sample	Pancreatic cancer (n = 30); Normal adjacent tissue (n = 30)	qRT-PCR	[175,176]
miR-127	High (N = 42)	Low (N = 42)	Snap-frozen sample	Pancreatic cancer (n = 42); Normal adjacent tissue (n = 42)	qRT-PCR	[158]
miR-181d	Low (N = 37)	High (N = 37)	Snap-frozen sample	Pancreatic cancer (n = 37); Normal adjacent tissue (n = 37)	qRT-PCR	[177]
miR-107	Low (N = 80)	High (N = 100)	Plasma	Pancreatic cancer (n = 100); Normal pancreas (n = 80)	qRT-PCR	[178]
miR-1290	Low (N = 267)	High (N = 167)	Plasma	Pancreatic cancer (n = 167); Normal pancreas (n = 267)	ddPCR	[179]

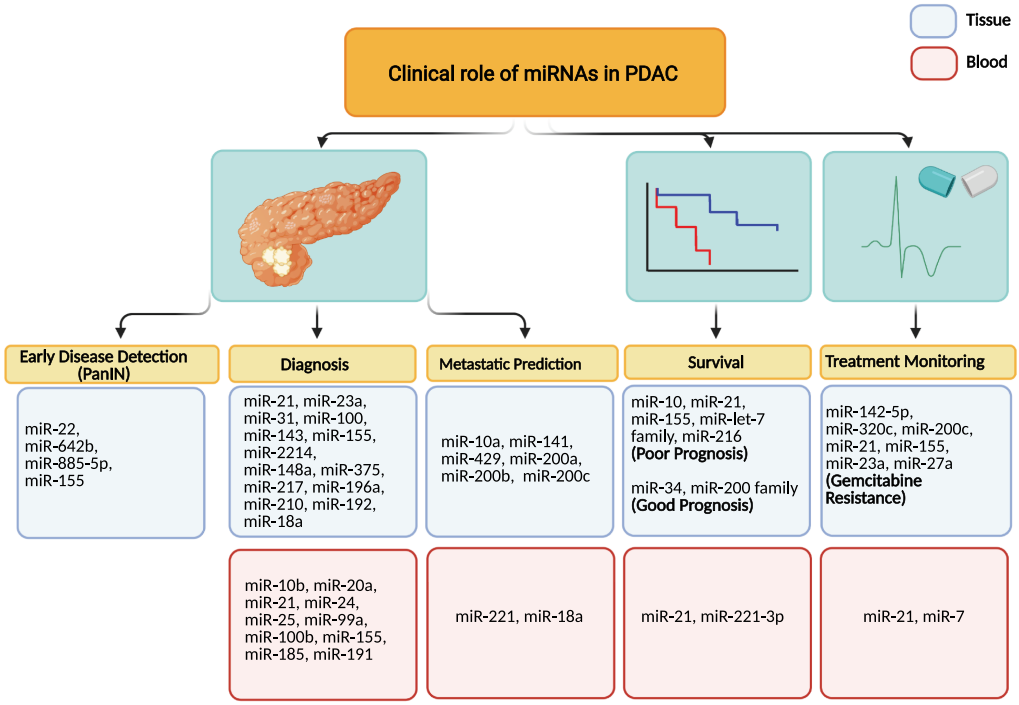


Figure 1. Clinical role of miRNAs in early PDAC detection, diagnosis, metastatic prediction, survival and treatment monitoring. The scheme shows different miRNAs, tissue and blood-derived, which could serve as biomarkers for discriminating the different stages of the disease, as well as for early diagnosis and metastasis prediction. Furthermore, some miRNAs could be associated with prognosis and monitoring of PDAC patients.

metastasis [38]. However, other microRNA, such as miR-10b, miR-200b, and miR-200 c, miR-218, miR-194, and miR-429 are also emerging biomarkers for metastasis in PDAC [39]. In blood, only miR-221 and miR-18a had been evaluated to be significantly associated with metastasis [40,41]. Unsurprisingly, all of the aforementioned miRNAs are also associated with patient's prognosis. The tissue expression of miR-10, miR-21, miR-155, miR-let-7 family, and miR-216 are known to predict an unfavourable prognosis, while expression of miR-34 and miR-200 family correlated with a better prognosis. For circulating miRNAs, only miR-21 and miR-221-3p have been consistently proven as indicators of poor prognosis. Despite the urgency in detecting metastasis in pancreatic cancer, there are only limited number of studies regarding the predictive value of microRNA in metastasis which limit their clinical validation and application.

MiR-21 in combination with miR-23a and miR-27a was also associated with more malignant PDAC phenotype and shorter overall survival after tumour resection [42]. In addition, expression of miR-21 determines PDAC response against gemcitabine with a lower expression correlating with better treatment outcome [43–45]. However, tumoural miR-21 overexpression emerged in a pooled meta-analysis assessing miRNAs as prognostic biomarkers in PDAC, independent of other clinicopathologic factors, including adjuvant chemotherapy use [44].

Differential expression of miRNA is not only observed in tissue samples but also in blood. For instance, miR-18a, miR-21, miR-22, miR-24, miR-25, miR-27a, miR-155, miR-185, miR-191, miR-196a, miR-642b and miR-885-5p were significantly upregulated in PDAC patients' blood plasma

[17,41]. Most importantly, blood-based miRNA profiling not only helps to differentiate PDAC patients from healthy individuals but also from other conditions that usually are considered as differential diagnosis such as acute or chronic pancreatitis and benign pancreatic tumours[32]. Another recent study showed that miR-486-5p and miR-938 could differentiate patients with PDAC from those who were healthy or had pancreatitis [46]. Additionally, circulating miRNAs can also be used as therapeutic biomarker. For example, downregulation of miR-181a-5p after FOLFIRINOX therapy correlates with better survival in PDAC but not in those who were treated with nab-paclitaxel and gemcitabine [47].

MiRNA dysregulation drives tumorigenesis through a close link with cellular signalling and metabolism. Several studies demonstrated that miR-21 enhanced PI3K/AKT and MAPK/ERK signalling that promote cell proliferation [48–50]. MiR-21 also suppresses the expression of phosphatase and tensin homolog (PTEN) and programmed cell death protein 4 (PDCD4) which facilitate cellular invasion induced by TGF-β signalling [51,52]. MiR-21 is also known to activate pancreatic stellate cells and cancer associated fibroblasts (CAF) to actively produce extracellular matrix proteins which contribute to its dense stroma [53,54]. On the other hand, miR-155 suppresses suppressor of cytokine signalling 1 (SOCS1) and MLH1 expression within cancer cells and enhance cancer invasion [55,56]. MiR-155 knock down is known to reduce membrane-type 1 matrix metalloproteinase (MT1-MMP), EGFR, and K-Ras expression in PDAC cell lines which led to lower proliferation rates and colony formation [57]. The functionality of other miRNAs has also been studied but seems less clear compared to miR-21 and miR-155.

Splicing factors and alternative splicing in PDAC

During malignant transformation, cancer cells experience aberrant splicing processes which result from mutations at the splice sites, mutations of splicing factors, and/or over/under expression of certain splicing factors [14]. Splicing deregulation could suppress protein expression by directing the inappropriately spliced mRNAs towards non-sense mediated decay or producing more active splice variants of oncogenic proteins [8,9,14]. The clinical application and implication of this process have also been studied both as therapeutic targets and biomarkers [8,15,58] and only recently received attention in PDAC.

Normal splicing is a post-transcriptional process where the introns are removed from primary transcripts, leaving only exons in the final transcript [14] (Figure 2a). In alternative splicing, exons can also be removed and different transcripts can be produced from a single gene [8]. RNA splicing occurs in the nucleus and is facilitated by splicing factors (SFs) which will assemble themselves in a sequential manner during the splicing process. The typical process is initiated by binding of a small nuclear ribonucleoprotein (snRNPs) to the primary transcript. Initially, U1 snRNP binds to the 5' splice site (5'SS) while U2 snRNP binds to a branch point at the other end of the intron. U1 and U2 then attract more snRNPs (U5, U4/U6) which then form a complete spliceosome and bend the intronic section, forming a lariat-like structure in which the 5' side of the intron is ligated to the branch point. Then, U4 is removed while the 5'SS site is hydrolysed [15]. In this process, the two extremities of the exons are held together by the spliceosome complex. Consecutively, the 3'SS is cut and the two exons are ligated, forming the final transcript that will be transported to the cytoplasm for translation [8,15,58].

Splicing is regulated by a wide array of splicing factors which bind to specific sites in primary transcripts [15,59]. The binding sites of those factors can be located in an exon or intron and can induce or repress the splicing process. The most important splicing factors and their binding sites are presented in Figure 2b. Splicing factors act early in the splicing process, facilitating snRNPs binding to primary transcripts [59]. Typically, SRSF2 binds the exonic splicing enhancer in exons flanking the intron and facilitate U1 and U2 binding. It connects to U1 by 70 K linker protein while its interaction with U2 is much more complex. It interacts with the U2 Small Nuclear RNA Auxiliary Factors U2AF1 and U2AF2, as well as with Zinc Finger CCCH-Type, RNA Binding Motif and Serine/Arginine Rich 2 (ZRSR2) and RNA-binding motif 10 (RBM10) in facilitating U2 binding. Additionally, SF3B1

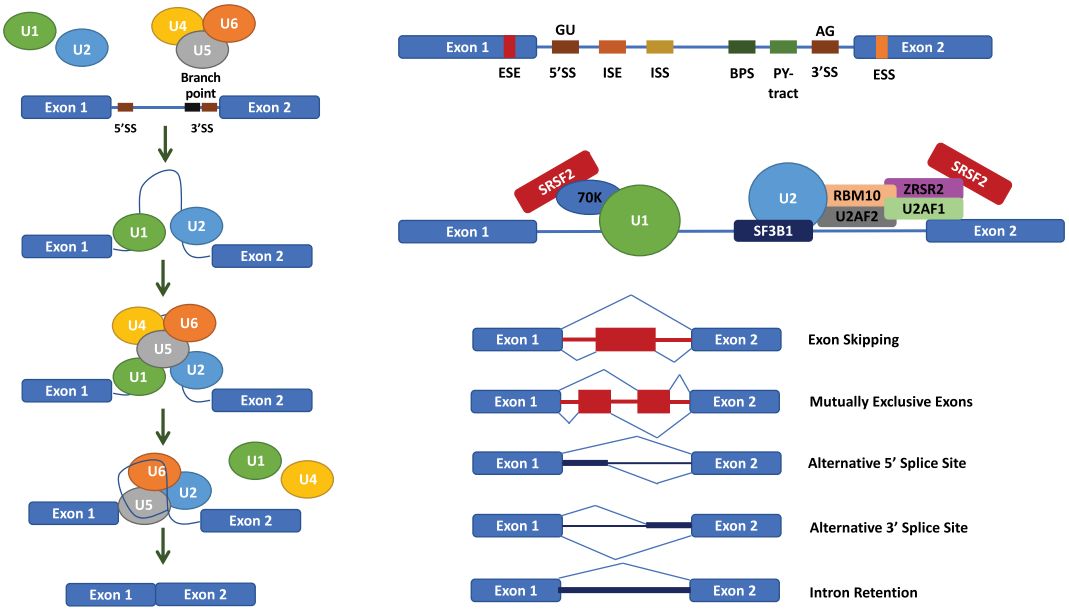


Figure 2. The mechanism of splicing mediated by splicing factors. The splicing is initiated by binding of U1 at 5′SS and U2 at 3′SS, bending the intron segment. Both SFs (U1 and U2) then recruit another SF which induced loop formation, cleaved the intron segment and ligated the exons [14,15]. Splicing is regulated by splicing factors which bind the primary transcript at several regulatory sites. Several essential regulatory sites and splicing factors are presented in the right part. SRSF2 is particularly important in initiating splicing by facilitating U1 and U2 binding to the primary transcript while SF3B1 mediates U2 binding to BPS. After U1 and U4 detached from spliceosome, only SRSF2 and SF3B1 remain in the complex while the other SFs detached [59]. ESE: Exonic splicing enhancer; ESS: Exonic splicing suppressor; ISE and ISS: Intronic splicing enhancer/suppressor; 5′SS and 3′SS: 5′ or 3′ splice site; BPS: Branch point site; Py-tract: Polypyrimidine-tract.

facilitates U2 binding by interacting with a branch point site [15]. After U1 and U4 dissociate from spliceosome, all of the splicing factors are also dissociated except SF3B1 which firmly binds to U2 and SRSF2[62].

RNA splicing and alternative splicing are crucial steps in protein expression and the isoforms of the proteins that are expressed by a certain gene are determined by these processes [8]. In cancer, these processes can be altered and this alteration can drive carcinogenesis [60]. In fact, in many types of cancer, splicing factors are either mutated or overexpressed, which strongly indicates an aberrant splicing process in cancer [8,15].

Mutations in splicing factors have been identified in several types of cancer as the driving force of carcinogenesis, most notably in haematologic cancers. SF mutations are detected in 78% of refractory anaemia with ringed sideroblasts and 60% of chronic myelomonocytic leukaemia (CMML) while it only happens in less than 5% in pancreatic, lung, breast, and head and neck cancer [15]. However, despite a lower mutational frequency, several SFs are over/under expressed in solid cancer including PDACs [8,14]. For example, SF3B1 and heterogeneous nuclear ribonucleoprotein K (HNRNPK) are consistently overexpressed in PDAC and are linked to an unfavourable prognosis [61–64]. Several important SFs in PDAC and their functions are summarized in Table 2.

Of note, in PDAC, SF3B1 and U2AF1 are the only known SFs with mutations and occur at a very low frequency [15,65]. However, these mutations are interesting because they can be targeted and induce synthetic lethality [66]. Furthermore, it appeared that PDAC relies on the normal form of SFB31, U2AF1, and RBM10 since patients with these mutations tend to have better prognosis compared to the wild types [9,14,15].

A more frequent form of splicing deregulation in PDAC consists in the overexpression of SFs [8]. Several SFs are upregulated in PDAC such as SF3B1, SRSF1, SRSF6, hnRNPK, Heterogeneous Nuclear Ribonucleoprotein A1 (hnRNPA1), and Polypyrimidine Tract Binding Protein 1 (PTBP1), while Rbfox2 tends to be downregulated . These splicing factors are thought to mediate many of PDACs unique characteristics, such as dense stromal, low immunogenicity, immune avoidance, as well as early metastasis, and invasion [9,64,65]. The pivotal role of splicing deregulation in PDAC was described by Wang et al. [8,58,59,61,64–66] who compared alternative splicing in PDAC to normal pancreatic tissue through Affymetrix exon array.

Table 2. Splicing factors in PDAC and their biological and clinical effects in preclinical studies.

Splicing Factors	Biological and Clinical Significance	Reference
	Upregulated in PDAC	[76,80]
	Upregulated by Myc	[76,80]
SRSF1	Promote resistance to gemcitabine	[76]
	Promote oncogenic splice variant of Bcl-xs, ΔRON and MCL-1s alternative splicing preferring their oncogenic variant	[77–79]
SRSF6	Increased proliferation and cellular transformation	[72]
	Prognostic factor for PDAC	[75]
SF3B1	Overexpression is associated with poor prognosis	[15,79,65]
	Mutated in 4% PDAC with mutation associated with better survival	[8,14]
	Important in branch point regulation and alters the splicing process of several oncogenes and tumour suppressor genes	
	Often downregulated as its control cellular proliferation	[97]
Rbfox2	Moderate upregulation in cancer tissue increased invasive potential	[98]
	May specify the mesenchymal tissue-specific splicing profiles both in normal and in cancer tissues	[8]
	Superfamily of RNA-binding proteins	[8]
HnRNPs	hnRNPA2B1 and hnRNPA1 altered Bcl-x alternative splicing and facilitate KRAS expression	[8,112]
	Higher expression associated with poor survival	[8]
HnRNPK	Wide range of effect including alteration in alternative splicing, gene transcription and RNA stability	[8,59,63,106,108,113]
	Altered PKM expression by favouring PKM2 and induces a Warburg effect	[70,124]
PTBP1	Upregulation of PTBP1 after chronic exposure to gemcitabine, conferring resistance against the drug	[70]

Alternative splicing tends to occur in genes encoding extracellular matrix (ECM), ECM-receptor interaction, and focal adhesion protein. In addition, pyruvate kinase and acyl-CoA synthetase long-chain family member 5 (ACSL5) were also present, which suggests that alternative splicing may have an impact on tumour metabolism.

Splicing deregulation drives pancreatic carcinogenesis by shifting the expression of pivotal oncogene and tumour suppressor proteins [5,8,14]. A clear example is the shifting of RON isoform expression of the tyrosine kinase receptor recepteur d'origine nantais (RON) [67,68]. Normally, RON has a low expression in normal pancreatic epithelial cells, but its expression increases gradually from low to high grade pancreatic intraepithelial neoplasia [68]. In PDAC, it is expressed in 69–96% of cases [68]. However, it is not only its higher expression that makes RON so important in PDAC; RON has indeed also several splicing alterations which lack exon 10, 11, or have 5 + 6 exon skipping [67]. These isoforms are constitutively active and, therefore, have more oncogenic potential compared to native isoform [67].

Another example is Pyruvate Kinase M2 (PKM2) whose expression has been observed in almost all types of tumours [69]. Normally, pancreatic epithelial cells express PKM1 instead of embryonic PKM2. This shift is mediated by PTBP1 which is also overexpressed in PDAC [70]. PTBP1 associates directly to intron 8 of PKM mRNA and induces alternative splicing. Therefore, the higher expression of PKM2 facilitates an oncogenic glycolytic metabolism and increases cancer cell resistance towards genotoxic drugs. This effect was confirmed by a knock-down study where PDAC cell lines with suppressed expression of PTBP1 or PKM2 are much more sensitive to Gemcitabine.

Deregulated splicing factors in PDAC and their interaction with miRNA expression

Several important splicing factor aberrations have been identified in PDACs and, most recently, their association with miRNA expression has been described in different cancer types. Although many known splicing factors are deregulated in cancer, PDAC is relatively unexplored and tends

to be limited to SF3B1. Overall, there are four splicing factors which have been studied in more detail and which will be discussed in detail.

SRSF6

SRSF6 is one of the most important splicing factors in PDAC and is often upregulated but not mutated [8,71]. SRSF6 is classified as an oncogenic splicing factor since it enhances cellular proliferation. Jensen et al. [72] reported that SRSF6 overexpression induces excessive keratinocyte hyperplasia in sensitized skin. Cohen-Eliav et al. [73] reported that SRSF6 overexpression increased proliferation rate of immortalized lung epithelial cells, transforming these into malignant cells. In addition, the human protein atlas considers SRSF6 as a potential prognostic marker in renal cancer, liver cancer, and PDAC [74]. Interestingly, and in contrast with the other two above-mentioned cancers, lower SRSF6 expression is correlated with poor prognosis in PDAC.

Keeping with these findings, Li et al. [75] showed that miR-193a-5p downregulates SRSF6, increasing the metastatic potential of PDAC cell lines. Apparently, SRSF6 downregulation was beneficial for cancer cells because it enhanced the invasive properties through the alteration of oxoglutarate dehydrogenase-like and ECM1 protein by alternative splicing. Thus, SRSF6 has a dual, apparently contradictive function since high expression of SRSF6 induces PDAC while a downregulation promotes tumour invasiveness. SRSF6 downregulation probably occurs late in the PDAC evolution, whereas at an early stage of carcinogenesis SRSF6 upregulation is preferred due to its beneficial effect on promoting cellular proliferation and survival. However, further studies in primary cells as well as in patient samples are needed to confirm this hypothesis.

SRSF1

SRSF1 is a well-characterized SR protein in cancer and one of SR proteins that is overexpressed in different cancers, including PDAC [8,59,71]. SRSF1 is a versatile protein, promoting carcinogenesis through several important mechanisms including increased proliferation rate and apoptosis resistance [76]. Furthermore, SRSF1 is a known splicing factor that influences the expression of Bcl-x, RON and MCL-1 isoform expression, changing their anti-apoptotic to pro-mitotic variants in cancer [77–79]. High expression of SRSF1 is induced by the Myc oncogene which is commonly upregulated in many cancers including PDAC [80].

Interestingly, besides its role in processing mRNA splicing, SRSF1 facilitates miRNA processing. SRSF1 is indeed involved in the final cleaving process mediated by Drosha, serving as an auxiliary factor [81]. The miR-7 family depends on SRSF1 for its maturation. However, miR-7 itself would suppress the expression of SRSF1, forming a negative feedback loop. Other miRNAs that depend on SRSF1 include miR-221, miR-222 and miR-17-92[86][82]. MiR-221/222 contribute to the progression of PDAC by increasing the expression of MMP-2 and MMP-9, increasing stromal remodelling and the invasive properties of the cancer cells [83]. For miR-17-92 which consists of four members, namely miR-17, miR-18, miR-19, and miR-92, an oncogenic effect was shown [84–87]. In PDAC, miR-19 actually promoted invadopodia and increased the invasiveness [88]. Therefore, it is important to further investigate the interaction between SRSF1 with miR-7, miR-221/222, and miR-17-92 to better understand the feedback loops that exist among them and also to assess why their overexpression favour carcinogenesis instead of tumour suppression.

SF3B1

SF3B1 is a well-known protein with a key role in PDAC [8,14,15]. It has the highest mutation rate among splicing factors and is also often overexpressed in PDAC [15]. Furthermore, this is one of the only three splicing factors that can be inhibited by small-molecule inhibitors so far, making its therapeutic potential higher than other SFs [89].

However, despite the wealth of information regarding the biological role of SF3B1 and its modulation, its relationship with miRNA is poorly understood. Their association has only been studied by Aslan et al. [90] in myelodisplastic syndrome in which they reported that the SF3B1

mutation was associated with global downregulation of tumour suppressor miRNAs from the let-7 family, especially miR-103a and miR-423. However, the mechanism of downregulation was not evaluated and there is still the possibility that this downregulation is not directly related to a SF3B1 mutation but might relate to other signalling aberrations.

Pianigiani et al. [91] demonstrated that there is a relationship between SF3B1 and splice site overlapping miRNAs (SO-miRNA). The precursors of these miRNAs are generated on the intron-exon junctions, from which the name 'splice-site' belongs. SF3B1 knockdown did not affect SO-miRNAs in HeLa and HaCAT cells, but increased the level of 52 SO-miRNA including miR-636, miR-6510-5p, miR-3614-3p, miR-3655, miR-3656, miR-4260, miR-5187-3p, miR-7109-5p, and miR-8069. Some of these miRNAs are classified as tumour suppressors [92–95]. However, these *in-vitro* results cannot be generalized to PDAC and gene silencing using siRNA is different than protein inhibition by a small molecule because there might be a different active site involved in miRNA processing than the inhibited site. Nevertheless, this finding suggests that SF3B1 modulation could have an additional beneficial effect by enhancing tumour suppressor miRNA expression. In addition, when these miRNAs are secreted to the extracellular compartment, they could also serve as biomarker for therapeutic monitoring.

Rbfox2

The RNA-binding Fox (Rbfox) proteins (Rbfox1, Rbfox2 and Rbfox3) constitute an important class of regulators of alternative splicing, and Rbfox2 (RBM9) can influence small and non-coding RNA in PDAC [8,59,96]. This RNA-binding protein is highly conserved in mammals [97], and it is different from Rbfox1 and Rbfox3, whose expression is limited to neuron and muscle cells. Rbfox2 is indeed widely expressed, especially in stem cells, haematopoietic stem cells, and embryos, where it regulates cellular proliferation [8]. In PDAC, Rbfox2 is downregulated, similar to other types of cancer [96]. These findings seem controversial because Rbfox2 is essential for cancer cell invasion, and the level of Rbfox2 increased moderately after the induction of EMT [98]. However, the same study showed that after the initial induction the levels of Rbfox2 were decreased. Notably, Rbfox2 is also subject to regulation by other proteins, and this might also momentary increase its expression [99]. Moreover, this mechanism might enable cancer cells to exploit the EMT promoting ability while evading excessive tumour suppressing effect by Rbfox2.

Rbfox2 is known to upregulate tumour suppressor miRNA, such as miR-20b and miR-107 while cleaving the oncomiR-21 [18–100]. The regulation of miRNA expression by Rbfox2 is mediated by direct binding to cognate sequences in miRNA or indirectly affects miRNA expression by altering Dicer expression. Of note, mutations in the DICER gene as well as in other components of the miRNA biogenesis pathway are not commonly detected in PDAC, and miRNA upregulation is more common than downregulation [100–102]. Additionally, several studies showed that miRNAs are broadly required for the development and maintenance of pancreatic cell lineages and play a role in carcinogenesis [103–105]. These findings suggest that miRNAs play a pivotal role in pancreatic tumorigenesis, and that loss of function mutations in the miRNA processing machinery are selected against during tumour evolution, but the impact of Dicer in later stages of pancreatic tumorigenesis or progression remains limited.

HnRNPs

Heterogeneous nuclear ribonucleoproteins (HnRNPs) are essential members of the RNA-binding proteins (RBPs) that act as regulators of alternative splicing, particularly, in linking the primary transcript with splicing machinery [106]. Several of its family members have been studied in relation to their role in carcinogenesis [106–108]. In PDAC, HnRNPA2B1 and HnRNPA1 are known for their role in tumour progression by shifting the Bcl-x isoform expression and facilitating KRAS expression, respectively [18,109]. In addition, HnRNPs are prognostic factors in PDAC with a higher expression associated with significantly shorter survival [108].

While there is no direct evidence of HnRNP and miRNA interaction in PDAC, their

interaction has been identified in ovarian cancer. Aguayo et al. [18] reported that HnRNPA1 suppresses miR-18a expression in docetaxel-resistant ovarian cancer cell lines. MiR-18a normally suppresses KRAS expression, but its downregulation enhanced KRAS expression and facilitated resistance to docetaxel [110,111]. MiR-18a has also been investigated in PDAC and elicited the same effect towards KRAS [112]. HnRNPA1 expression was also suppressed by miR-15a-5p and miR-25-3p [18]. These two miRNAs are also known as tumour suppressor miRNAs in PDAC [7]. Therefore, the same molecular mechanism might exist in PDAC, and inhibiting or blocking HnRNPs could be explored as a new way to fight chemoresistance in this disease.

Another member of HnRNPs that was recently investigated regarding its role in PDAC is HnRNPK [6,8,14,106]. Much of the biology of HnRNPKs is still under investigation because these proteins are not only involved in RNA splicing but also in DNA transcription and RNA stability [113]. They are also responsible for the downregulation of some tumour suppressor genes in PDAC [63,64]. Remarkably, HnRNPKs interact with miR-223, an oncomiR that enhances cell proliferation and migration [63]. These effects have been attributed to downregulation to miR-223 targets FBXW7 and PDS5B, two tumour suppressor proteins which inhibit cellular migration and induce apoptosis [63,114]. A similar finding on the importance of miR-223 was also found in pancreatic cancer cells when using the naturally occurring isoflavonic phytoestrogen genistein that inhibited miR-223 expression which in turn enhanced FBXW7 expression [115]. These effects resulted in inhibition of cell growth and induction of apoptosis.

PTBP1

PTBP1 has been investigated for its role in PDAC metabolism [70,108]. The expression of PTBP1 was increased in two Gemcitabine resistant PDAC cell lines (PANC-1 and Pt45P1) where it modulated alternative splicing alteration of PKM, resulting in overexpression of the cancer-related PKM2 isoform, whose high expression also correlated with worse prognosis in PDAC patients [70]. PTBP1 is also considered a prognostic factor and a potential therapeutic target due to its role in enhancing PDAC metabolism [108].

The only proven miRNA that directly interacts with PTBP1 in PDAC is miR-124, which directly downregulates PTBP1 mRNA and shifts PKM isoform expression from PKM2 to PKM1. The importance of miR-124 and PTBP1 interaction was shown by ectopic expression of miR-124 or administration of PTBP1 siRNA which increased sensitivity to gemcitabine and relieved autophagy in gemcitabine resistant PDAC cell lines [116]. However, in PDAC, miR-124 is mostly downregulated which facilitates increased PTBP1 expression, favouring Warburg effect [117,118]. In neural differentiation, Yeom et al. [119] observed that PTBP1 could repress miR-124 maturation by directly binding to pri-miR-124 and blocked transcript cleavage by DROSHA. Therefore, we can assume that a low expression of miR-124 could also result from increased expression of PTBP1 and this potential feedback loop adds to the complexity of splicing factors-miRNA interaction in cancer.

Another miRNA known to interact with PTBP1 is miR-133b [119]. Although there are no data regarding their interaction in PDAC, miR-133b is downregulated in PDAC and has been considered as a tumour suppressor miRNA based on findings in other cancer types [120–123]. In colorectal cancer, miR-133b silenced PTBP1 expression and inhibited the Warburg effect by promoting the expression of the PKM1 isoform [124]. Due to its low expression in PDAC, miR-133b could also exert a similar effect in PDAC. Despite the limited direct evidence, there is a strong indication of interaction between splicing deregulation and miRNA in PDAC. A summary of relevant splicing factors and the miRNAs that interact with each other as well as their main biological effects is presented in Table 3. Remarkably, further studies exploring this field of research are now extremely timely since splicing inhibitors are becoming available as novel anticancer drugs and could offer a new therapeutic strategy for PDAC.

5 miRNA profiling methods

Accurate detection and quantification of miRNAs represent a major challenge due to the small size of

miRNAs (approximately 22 nucleotides), the high sequence homology among members of the same family and the low abundance in biofluids. Currently, miRNAs profiling is a growing field of study, although conventional methods for detecting miRNAs still remain the gold-standards used to confirm the results of new detection techniques [125].

Northern blot is a widely used historical method to measure the expression of miRNAs ranging from the primitive miRNA to the mature form. It is based on molecular hybridization and gel electrophoresis and is able to simultaneously determine the size of miRNAs. However, Northern blot has several disadvantages: it is a time-consuming technique, requires large amounts of samples and reagents, with low sensitivity (pM-nM range) and low throughput [126].

Current miRNA detection strategies include reverse transcription-quantitative polymerase chain reaction (RT-qPCR), which is so far the undeniably gold-standard method for routine testing, especially for diagnostic purposes. It is commonly used to detect miRNAs at any stage of maturation, but does not allow the identification of new miRNAs. RT-qPCR is less time-consuming technique than Northern blotting and displays higher sensitivity, specificity and reproducibility than Northern blot. In addition, it converts small miRNA sequences into longer sequences by adding a poly(A) tail (poly(A)-tailed RT-qPCR) or a stem-loop structure (stem-loop RT-qPCR)

Table 3. Relevant splicing factors and miRNAs affecting key aggressive biological features of PDAC in preclinical studies.

Main Effect	Splicing Factor	Associated miRNA	Interaction	Biological Impact	Reference
PROLIFERATION					
	SF3B1	miR-636, miR-6510-5p, miR-3614-3p, miR-3655, miR-3656, miR-4260, miR-5187-3p, miR-7109-5p, miR-8069, miR-155-3p, miR-148a-3p, miR-98-5p, and miR-21-3p	Upregulation miR-636, miR-6510-5p, miR-3614-3p, miR-3655, miR-3656, miR-4260, miR-5187-3p, miR-7109-5p, and miR-8069 Downregulation miR-155-3p, miR-148a-3p, miR-98-5p, and miR-21-3p	Decreased cellular proliferation Enhanced keratinocyte differentiation	[43,91]
	HnRNPs	miR-18a, miR-15a-5p, miR-25-3p	Downregulation of miR-18a by HnRNPA1 Suppression of HnRNPA1 by miR-15a-5p and miR-25-3p	Increased KRAS activation resulted from miR-18a downregulation	[18]
		miR-223	Increased expression of miR-223 by HnRNPK	Increased cancer cell proliferation by FBXW7 and PDS5B suppression	[63]
EPITHELIAL-TO-MESENCHYMAL TRANSITION AND METASTASIS					
	SRSF1	miR-7 family, miR-17, miR-18	Facilitate miR-7 family biosynthesis Increased miR-221/222 expression	Increased invasion and metastasis through MMP-2 and −9 upregulation	[81,83]
	SRSF1	miR-19, miR-92	Facilitate miR-17-92 family biosynthesis	Increased invasion and metastasis through MMP-2 and −9 upregulation	[82,83]
	SRSF6	miR-193a-5p	SRSF6 downregulation by miR-193a-5p	Facilitate metastasis by alteration in oxoglutarate dehydrogenase-like (OGDHL) and extracellular matrix protein 1 (ECM1) alternative splicing	[75]
	Rbfox2	miR-20b, miR-21	Upregulation of miR-20b Suppressing miR-21 expression	Considered as anti-cancer splicing factors; suppressing cellular proliferation at normal tissue	[180]
		miR-107	Upregulation of miR-107	Considered as anti-cancer splicing factors; suppressing cellular proliferation at normal tissue Enhancing EMT in PDAC when moderately increased in PDAC	[181]
TUMOUR METABOLISM					
	PTBP1	miR-124	In neuron: PTBP1 suppress miR-124 cleavage (not confirmed in cancer)	Enhanced resistance against gemcitabine	[116,119]
		miR-124, miR-133b	PTBP1 downregulation by miR-124 and miR-133b	Altered cancer metabolism favouring Warburg effect by promoting PKM2 expression	[124]

overcoming the primer design limitation [127]. An innovation is represented by the ddPCR (droplet digital PCR) which offers greater performance, improved sensitivity, and accuracy as it allows for absolute quantification of miRNAs without the need for a reference gene [128].

PCR techniques cannot detect nucleotide sequences in cells and tissue sections, while in situ hybridization (ISH) can visualize miRNAs within cells and can determine the spatiotemporal expression of miRNAs, elucidating their biological role as well as their pathologic involvement in numerous diseases [129]. This technique is labour intensive and is limited by its low-throughput nature but the recent development of directly labelled fluorescence probes and multiplexed miRNA ISH methods allowed to detect multiple miRNAs per reaction.

Microarray is a hybridization-based method suitable for relative quantification. Locked nucleic acids (LNAs) can be incorporated into capture probes to normalize the melting temperature (Tm) whose variance is related to miRNA GC content [128,130]. The strength of this method is the multiplexed detection of multiple miRNAs in a single reaction, although it cannot discriminate between miRNA variants and has poor sensitivity compared to RNA-seq because it lacks the amplification step. On the other hand, microarray assays are fast, expensive and high-throughput [131].

Finally, Next-Generation Sequencing (NGS) is a highly accurate miRNA profiling technique that can simultaneously measure expression level and sequence changes, as well as detect unknown miRNAs. It should be noted that NGS has the highest multiplexing capability as specific primers are not required for each targeted miRNA detection [125,132]. Drawbacks to NGS include time-consuming for converting a sample into a library for sequencing, expensive analyses due to sophisticated software and qualified personnel for data analysis and it is not a fully automated technique as well [126].

Since multiplexing capability plays a crucial role in miRNAs detection, in addition to the above-mentioned multiplexing approaches, it is worth citing the suspension arrays (i.e., on-particle), which represent promising emerging methods for highly multiplex analysis of complex samples due to the versatility of the encoded microspheres used in conjunction with flow cytometry [133]. Furthermore, Rondelez et al. recently reported an isothermal amplification mechanism for multiplex and digital detection of miRNAs using the rational building of a molecular circuit that suppresses non-specific amplification due to cross-talk reactions [134]. In conclusion, extensive efforts have been made so far to develop efficient and sensitive methods for miRNA detection, but there still remains a need for a standardized method that should be highly sensitive, specific and multiplexable.

Predicting the effect of splicing modulation and its effect towards miRNA profile of PDAC

Splicing modulation is a new emerging therapeutic approach that had been tested in several types of cancer either pre-clinically or clinically [15,66,89,135–137]. Splicing modulation is promising because of its high potency to induce apoptosis and suppress cellular migration. Cancer cells harbouring mutations in genes encoding splicing factors are the most promising targets [15,89,136,137]. However, splicing modulation could also be applied in cancer cells with splicing factors overexpression [9,89,135].

The important role of splicing deregulation in PDAC carcinogenesis and the potency of several splicing modulators might increase the potential application of splicing inhibitors in PDAC. However, there have not yet been studies evaluating the efficacy of splicing inhibitors/modulators in PDAC. Preclinical studies have demonstrated that the SF3B1 inhibitors pladienolide B and E7107 were effective in gastric cancer, cervical cancer, and peritoneal mesothelioma [135,138–140]. In particular, pladienolide B has high efficacy in gastric cancer with complete tumour elimination in SCID mice in just 2 weeks and has an IC_{50} in the nanomolar range. Similar findings were observed in peritoneal mesothelioma where pladienolide B and E7107 inhibited cell proliferation and migration [135]. Remarkably, *in vivo* treatment with E7107 resulted in complete regression of peritoneal tumours in the second week. SF3B1 inhibition also showed similar efficacy in cervical cancer and cutaneous squamous cell carcinoma but, in

these tumours they apparently showed more efficacy towards cells with mutated p53[140].

Unfortunately, clinical trials with splicing modulators have been limited by toxicity. Indeed, in a phase I pharmacokinetic and pharmacodynamic study of E7107 in advanced solid tumours, when using doses above 4.3 mg/m^2, several patients suffered from gastrointestinal side effects, such as diarrhoea, vomiting, dehydration, and in two cases there was vision loss [141].

H3B-8800 is another SF3B1 inhibitor and entered phase I clinical trial in 2016, with a focus on patients with MDS, AML and CMML (NCT02841540). Initial results revealed dose-dependent target engagement, a predictable pharmacokinetic profile and a favourable safety profile, even with prolonged dosing. Although objective therapeutic responses have not been achieved to date, 14% of patients had reduced requirements for red blood cell or platelet transfusions [142].

A number of other drugs targeting splicing factors have shown encouraging preclinical effects in mouse models of cancer, such as inhibitors of SRPK and CLK protein kinases that phosphorylate SR proteins and thereby inhibit angiogenesis by inducing changes in the alternative splicing of VEGF [143,144]. Other splicing inhibitors targeting a variety of spliceosomal components also reduce cancer cell proliferation in vitro [145–148], but their effects in animal models of cancer are not yet known.

However, toxicity may be prevented by the use of a lower dose of splicing modulators, and the risk of reduced efficacy can be avoided by rationale combinations with different antitumor strategies, including modulation of selected miRNAs.

There are no data yet on the potential effect of splicing inhibitors on miRNA in PDAC. The most plausible candidates as therapeutic targets are PTBP1 and HnRNPK because their role has been already established in PDAC [63,70]. Moreover, SF3B1 has the advantage as a therapeutic target due to the availability of small-molecule inhibitors [89,137,139,140,149]. Of note, SF3B1 inhibition resulted in upregulation of tumour suppressor SO-miRNAs in a cervical cancer cell line [91]. Therefore, similar studies should be performed to demonstrate this effect in PDAC cell lines.

The effect of splicing inhibitors targeting those three splicing factors might however be predicted using available data. Calabreta and colleagues [70] provided initial evidence that targeting the splicing factor PTBP1 in gemcitabine resistant PDAC cell line by siRNA shifted PKM isoform expression towards PKM1 which was accompanied by increased sensitivity towards gemcitabine and an enhanced level of cleaved caspase-3. Li et al. [116] studied the long non-coding ROR in PDAC and found that PTBP1 was the target of tumour suppressor miR-124 which could effectively block its expression. However, in PDAC, long non-coding ROR acts as sponge that binds miR-124, preventing it to regulate PTBP1 expression and increasing PKM2 expression. This study suggested that miR-124 could be used as a marker for Warburg effect in PDAC as well as a therapeutic agent candidate to target PTBP1 in gemcitabine resistant PDAC. However, the other targets of miR-124 should be elucidated to minimize unfavourable off-target effects.

Another potential SF target candidate is HnRNPK which is known for its role in enhancing cancer cellular proliferation, invasion and metastasis in PDAC [63]. HnRNPK is associated with miR-223 which suppressed FBXW7 as previously described. However, the sister chromatid cohesion protein PDS5 homolog B (PDS5B) is another important target of miR-223. The downregulation of miR-223 led to increased expression of PDS5B which resulted in inhibition of cellular proliferation and migration [114].

Inhibition of SF3B1 could possibly be effective and may produce the most pronounced miRNA profile changes in PDAC. In cervical cancer, SF3B1 inhibition resulted in an increase of several tumour suppressor miRNA, most notably miR-636, miR-6510-5p, miR-3614-3p, miR-3655, miR-3656, miR-4260, miR-5187-3p, miR-7109-5p, and miR-8069 [96]. In addition, four miRNAs were downregulated, namely miR-155-3p, miR-148a-3p, miR-98-5p, and miR-21-3p. Apparently, SF3B1 inhibition can suppress the expression of miR-155 and miR-21 which play important roles in PDAC. However, this should be further investigated in PDAC preclinical models.

The effect of upregulation of tumour suppressor miRNAs or downregulation of oncogenic miRNAs is expected to have a wide impact [25]. A summary of potential effects of splicing modulation on the miRNA profile in PDAC as well as their biological effects is depicted in Figure 3. For example, miR-21 and miR-155 have many targets that are involved in carcinogenesis and metastasis [11,12,50,53,55,56]. Suppression of these oncogenic miRNAs can thus potentially lead to tumour suppression and inhibition of metastasis. However, these miRNAs can also serve as potential biomarkers of tumour progression or response to treatment, and could improve the clinical management of PDAC patients by monitoring the modulation of these miRNAs in samples that can be collected during treatment/follow-up, such as in liquid biopsy studies.

Bioinformatic tools to predict the effect of splicing modulation towards miRNA profiles

Bioinformatics uses advanced computing, mathematics and biological knowledge to store, manage, analyse and get insights into biological data. In recent years, there has been a boom of publicly available computational tools, online data analysis modules, biological data repositories, and bioinformatics workflow management systems [150]. In order to assess how splicing modulation can affect miRNA profiles, alternative splicing detection tools such as rMATS [151], SUPPA2 [152] or MISO [153] can first detect differential splicing between conditions, after which splicing motif analysis tools like MEME [154] or RNAContext [155] can use these splicing motifs to identify regulators of alternatively spliced junctions. Lastly, potential miRNA targets of splicing

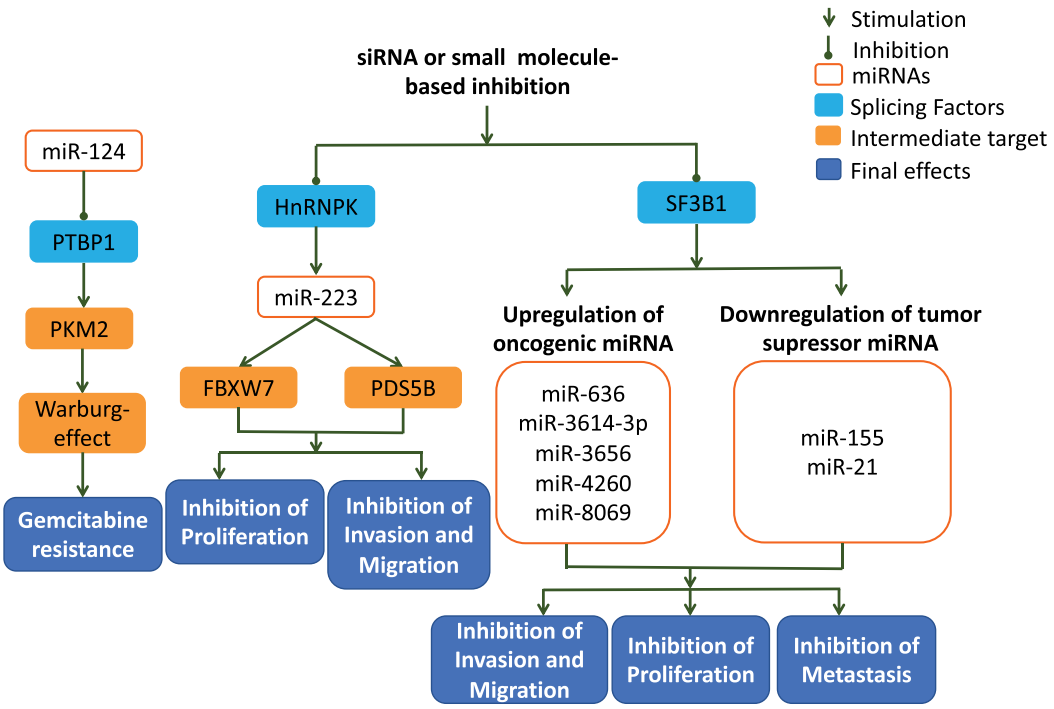


Figure 3. The interaction of relevant splicing factors in PDAC and their associated miRNA. PTBP1 and HnRNPK are considered as the relevant targets in pancreatic cancer and have demonstrated their interaction with miRNAs in PDAC (miR-124 and miR-223, respectively). Despite lack of evidence in PDAC, cervical cancer experiment demonstrated that SF3B1 inhibition resulted in extensive change in miRNA expression and potentially brings more profound effects than PTBP1 and HnRNPK [63,70,89,91].

modulation can be detected using miRNA-target databases such as mirTarBase [156] where experimentally validated miRNA-target interactions are curated, or mirDB [157] where the predictive algorithm MirTarget is used to analyse thousands of miRNA-target interactions from high-throughput experiments. An example of analysis pipeline to identify splicing factors modulated by miRNAs is reported in Figure 4.

The first step is the identification of alternative spliced (AS) events. While there is a plethora of available tools, no current tool can be regarded as the golden standard and the matter of choice strictly depends on the research question and familiarity. Research into which tools are superior is currently incomplete. One example of a tool that detects differential alternative spliced RNA transcripts is MISO, published in 2010 [157]. This statistical model estimates expression of alternatively spliced exons and isoforms using mapped reads as input format. MISO then uses Bayesian inference to compute the probability that an RNA-seq read originated from a particular isoform. Despite being the most cited and used tool for alternative spliced differential analysis, it has no longer been maintained since its publication, and it has high computational time. rMATS [151] is a more recent and often-cited tool used in differential splicing analysis, it analyses replicates and includes a function to handle paired and unpaired replicates. Another common tool recently published is SUPPA2[156]. This type of algorithm requires two biological replicates because it accounts for biological variability, which is important for the reliability of the estimations that are drawn from the data. However, it can work with multiple conditions and includes the possibility to perform hierarchical clustering on differentially spliced events to identify common regulatory mechanisms. Other tools capable of detecting AS and compare AS patterns between sample groups are DEXSEq, SplicingCompass, Altanalyze, BitSeq, EBSeq, and Cuffdiff2 whose performances are extensively reviewed in Lahat and Grellsheid [158].

Once differentially expressed AS events are identified, prediction of splicing factors is the second step of the analysis. Motif analysis tools may be used to identify the direct regulators of alternative spliced junctions. For example, MEME suite [154] provides a unified portal of

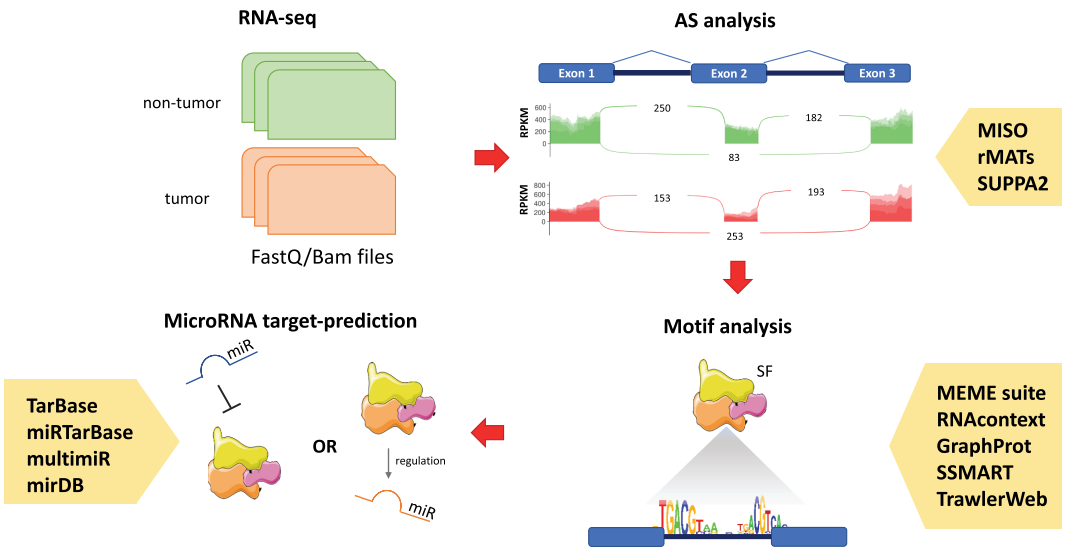


Figure 4. Bioinformatic pipeline for splicing factor and miRNA modulation discovery with RNA-seq data. Raw data or mapped BAM files are used as input for alternative splicing (AS) tools such as MISO, rMATs and SUPPA2 to identify differential alternative splicing events. Next, a motif analysis is performed to identify the splicing factor (SF) specific for that RNA isoform. Lastly, miRNA-target databases are used to retrieve possible miRNA targeting SF. Additional miRNA profile can be useful to detect miRNA modulation through SFs.

online discovery tools for DNA binding sites and protein interaction domains. MEME is an online web-based module that includes three sequence scanning algorithms that allow to scan different DNA and protein databases. Next, transcription factor motif can be further analysed for putative functions using GOMo tool [159,160]. Previous tools such as RNAcontext [155] and GraphProt [143] work with classification and regression model settings and are not capable of *de novo* sequence-structure motifs. Other tools for motif analysis recently published are SSMART [161] and TrawlerWeb [162] which is a web version of the previous published standalone tool. Trawlerweb is currently the fastest online de novo motif discovery tool and it displays resulting scores allowing the user to prioritize the choice for validation experiments. Validation analysis, such as *in vitro*/*in vivo* binding assays, cross-linking immunoprecipitation sequencing (CLIP-seq), minigene splicing reporter assays (*invitro*) or antisense oligonucleotides which block splicing factor-binding sites are needed to validate results from the (splicing) motif analysis.

The last step of the analysis is the identification of putative miRNAs that are targeting the splicing factors based on previously performed motif analysis. Nowadays there is a plethora of miRNAs-target databases available on the web, such as miRTarBase [156], mirDB [157], miRBase [163] and TarBase [164]. In addition, a new R package multimiR [165] includes a compilation of around 50 million records in human and mouse from 14 different databases and it expands on miRNAs involved in drug response and disease annotation.

However, an integrative analysis with miRNA profile can be useful to detect miRNA modulation (e.g., inhibition) through the predicted SFs. Nowadays, there are several ways to analyse miRNA-seq profiles, and here we describe the general downstream analysis pipeline with the most commonly used tools.

After trimming and quality check, the resulting reads are aligned to a reference database containing miRNA sequences. miRbase [163] is the primary database of published miRNA sequences which is often used for miRNA mapping. A broader database of small-RNA and miRNA sequences is mirGeneDB 2.0 [166] resulting in more precise annotation while avoiding misleading miRNA annotation from other types of small-RNAs. The read sequences are mapped to reference databases through mapping tools. There is an increasing number of mapping tools for small-RNA sequences and the most used are: miRanalyzer [167], miRDeep2 [168] and sRNAbench [169]. All these tools rely on Bowtie algorithm [170] (allowing mismatches and improving speed of alignment).

sRNAbench and its downstream analysis tool sRNAtoolbox includes an automatic processing of the five most used library preparation protocols (including new reference genomes from Ensembl, NCBI and MirGeneDB), a consensus differential expression analysis, target prediction, analysis of unmapped reads, batch mode to profile several samples at once with the same set of parameters and improved visualization and mapping statistics. This also enables users with a 'non-bioinformatics' background to analyse small-RNA high-throughput data from raw fastq files (standard output files from sequencing machines) to post-processed data for differential analysis and miRNA-target prediction.

The ever-expanding field of bioinformatic studies and the enormous availability of wet-lab data has given rise to several predictive models that are extremely useful for target prediction and to prioritize experimental validation targets.

Conclusions and Future Perspectives

The interaction of miRNAs and splicing deregulation is an understudied field, but evidence of their close interconnection is increasing. Currently, the application of miRNAs is focused on their role as biomarker while splicing inhibitors are under investigation as a novel therapeutic strategy.

Increasing evidence shows that splicing deregulation resulting from mutation or overexpression can produce a pronounced aberration in miRNA expression in different cancer types, including PDAC. For instance, the upregulation of tumour suppressing miRNAs may mediate an anti-cancer effect of splicing modulation as was shown by the inhibition of PTBP1 and SF3B1 in cervical cancer. Potentially similar inhibitory effects and the impact of other SFs on cancer progression and miRNA profile still need to be investigated in PDAC. Moreover, specific

miRNAs could be used as a target to downregulate specific SFs and also for combined therapeutic approaches.

Remarkably, novel bioinformatics tools are providing extensive data that can be used to deepen our knowledge in the biological effects of the interplay between splicing and miRNAs, as well as several predictive models for target prediction in order to prioritize future experimental and clinical validation. In-depth analysis of PDAC aberrant splicing patterns associated with miRNA profiling may indeed further provide mechanistic insight to successfully target key PDAC drivers. Targeting aberrant splicing and the reciprocal interaction with deregulated miRNA could therefore provide more effective therapeutic approach to combat the complex biology of PDAC and its chemoresistant features.

Authors' contribution

D.I.S, G.M., I.B.M., O.R., G.J.P, E.G. wrote the manuscript; D. I.S., O.R. and G.M. designed figures and tables; G.J.P, E.G., P. D. and S.C. extensively revised the manuscript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the CCA Foundation 2012, 2015 and 2018 grants (G.J.P., E.G.,); KWF Dutch Cancer Society grants [KWF project#19571] (E.G.); AIRC/IG-grant (E.G.); Indonesian Endowment Fund Scholarship (I.G.P.S.); KWF Kankerbestrijding [19571]; Indonesian endowment fund scholarship;AIRC/IG-grant;

ORCID

I Gede Putu Supadmanaba http://orcid.org/0000-0003-1577-0465
Giulia Mantini http://orcid.org/0000-0003-2947-9978
Ornella Randazzo http://orcid.org/0000-0003-1728-1673
Godefridus J. Peters http://orcid.org/0000-0002-5447-2877
Elisa Giovannetti http://orcid.org/0000-0002-7565-7504

References

[1] Kleeff J, Korc M, Apte M, et al. Pancreatic cancer. Nat Rev Dis Primers. 2016;2:16022.
[2] The Netherland Cancer Registry [Internet]. Dutch Cancer Figure: Pancreatic Cancer2018 [cited 2019 Feb 22]; Available from: https://www.cijfersoverkanker.nl/selecties/Dataset_1/img5c8269c859485
[3] McGuigan A, Kelly P, Turkington RC, et al. Pancreatic cancer: a review of clinical diagnosis, epidemiology, treatment and outcomes. World J Gastroenterol. 2018;24(43):4846–4861.
[4] Fahrmann JF, Bantis LE, Capello M, et al. A Plasma-Derived Protein-Metabolite Multiplexed Panel for Early-Stage Pancreatic Cancer. J Natl Cancer Inst. 2019;111(4):372–379.
[5] Orth M, Metzger P, Gerum S, et al. Pancreatic ductal adenocarcinoma: biological hallmarks, current status, and future perspectives of combined modality treatment approaches. Radiat Oncol. 2019;14:141.
[6] Yang C, Wu Q, Huang K, et al. Genome-Wide Profiling Reveals the Landscape of Prognostic Alternative Splicing Signatures in Pancreatic Ductal Adenocarcinoma. Front Oncol. 2019;9:511.
[7] Słotwiński R, Lech G, Słotwińska SM. MicroRNAs in pancreatic cancer diagnosis and therapy. Cent Eur J Immunol. 2018;43(3):314–324.
[8] Anczuków O, Krainer AR. Splicing-factor alterations in cancers. RNA. 2016;22(9):1285–1301.
[9] Wang J, Dumartin L, Mafficini A, et al. Splice variants as novel targets in pancreatic ductal adenocarcinoma. Sci Rep. 2017;7(1):2980.
[10] Hu G, Tao F, Wang W, et al. Prognostic value of microRNA-21 in pancreatic ductal adenocarcinoma: a meta-analysis. World J Surg Oncol. 2016;14(1):82.
[11] Sicard F, Gayral M, Lulka H, et al. Targeting miR-21 for the therapy of pancreatic cancer. Mol Ther. 2013;21(5):986–994.
[12] Feng Y-H, Tsao C-J. Emerging role of microRNA-21 in cancer. Biomed Rep. 2016;5(4):395–402.
[13] Daoud AZ, Mulholland EJ, Cole G, et al. MicroRNAs in Pancreatic Cancer: biomarkers, prognostic, and therapeutic modulators. BMC Cancer. 2019;19(1):1130.
[14] Escobar-Hoyos L, Knorr K, Abdel-Wahab A-WO. Aberrant RNA Splicing in Cancer. Annu. Rev. Cancer Biol. 2019;3(1):167–185.
[15] Dvinge H, Kim E, Abdel-Wahab O, et al. RNA splicing factors as oncoproteins and tumour suppressors. Nat Rev Cancer. 2016;16(7):413.
[16] Kahn T, Bosch J, Giovannetti E, et al. Effect of sodium nitrate loading on electrolyte transport by the renal tubule. The American Journal of physiology. 1975;229 (3):54714.
[17] Ali S, Almhanna K, Chen W, et al. Differentially expressed miRNAs in the plasma may provide a molecular signature for aggressive pancreatic cancer. Am J Transl Res. 2010;3(1):28–47.
[18] Rodriguez-Aguayo C, Monroig P, Del C, et al. Regulation of hnRNPA1 by microRNAs controls the miR-18a–K-RAS axis in chemotherapy-resistant ovarian cancer. Cell Discov. 2017;3:17029.

[19] Krol J, Loedige I, Filipowicz W. The widespread regulation of microRNA biogenesis, function and decay. Nat Rev Genet. 2010;11(9):597–610.

[20] Juzenas S, Venkatesh G, Hübenthal M, et al. A comprehensive, cell specific microRNA catalogue of human peripheral blood. Nucleic Acids Res. 2017;45 (16):9290–9301.

[21] Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell. 2004;116(2):281–297.

[22] Winter J, Jung S, Keller S, et al. Many roads to maturity: microRNA biogenesis pathways and their regulation. Nat Cell Biol. 2009;11(3):228–234.

[23] Zhang F, Wang D. The Pattern of microRNA Binding Site Distribution. Genes (Basel). 2017;8(11):296.

[24] Hogg DR, Harries LW. Human genetic variation and its effect on miRNA biogenesis, activity and function. Biochem Soc Trans. 2014;42(4):1184–1189.

[25] Peng Y, Croce CM. The role of MicroRNAs in human cancer. Signal Transduct Target Ther. 2016;1(1):15004.

[26] Poy MN, Eliasson L, Krutzfeldt J, et al. A pancreatic islet-specific microRNA regulates insulin secretion. Nature. 2004;432(7014):226–230.

[27] Olson P, Lu J, Zhang H, et al. MicroRNA dynamics in the stages of tumorigenesis correlate with hallmark capabilities of cancer. Genes Dev. 2009;23 (18):2152–2165.

[28] Schultz NA, Werner J, Willenbrock H, et al. MicroRNA expression profiles associated with pancreatic adenocarcinoma and ampullary adenocarcinoma. Mod Pathol. 2012;25(12):1609–1622.

[29] Hong TH, Park IY. MicroRNA expression profiling of diagnostic needle aspirates from surgical pancreatic cancer specimens. Ann Surg Treat Res. 2014;87 (6):290–297.

[30] Calatayud D, Dehlendorff C, Boisen MK, et al. Tissue MicroRNA profiles as diagnostic and prognostic biomarkers in patients with resectable pancreatic ductal adenocarcinoma and periampullary cancers. Biomark Res. 2017;5(1):8.

[31] Papaconstantinou IG, Manta A, Gazouli M, et al. Expression of microRNAs in patients with pancreatic cancer and its prognostic significance. Pancreas. 2013;42(1):67–71.

[32] Rawat M, Kadian K, Gupta Y, et al. MicroRNA in Pancreatic Cancer: from Biology to Therapeutic Potential. Genes (Basel). 2019;10(10):752.

[33] Frampton AE, Krell J, Prado MM, et al. Prospective validation of microRNA signatures for detecting pancreatic malignant transformation in endoscopic-ultrasound guided fine-needle aspiration biopsies. Oncotarget. 2016;7(19):28556–28569.

[34] Du Rieu MC, Torrisani J, Selves J, et al. MicroRNA-21 Is Induced Early in Pancreatic Ductal Adenocarcinoma Precursor Lesions. Clin Chem. 2010;56(4):603–612.

[35] Ryu JK, Hong S-M, Karikari CA, et al. Aberrant MicroRNA-155 Expression Is an Early Event in the Multistep Progression of Pancreatic Adenocarcinoma. Pancreatology. 2010;10(1):66–73.

[36] Yu J, Li A, Hong S-M, et al. MicroRNA Alterations of Pancreatic Intraepithelial Neoplasias. Clin Cancer Res. 2012;18(4):981–992.

[37] Caponi S, Funel N, Frampton AE, et al. The good, the bad and the ugly: a tale of miR-101, miR-21 and miR-155 in pancreatic intraductal papillary mucinous neoplasms. Ann Oncol. 2013;24(3):734–741.

[38] Wei X, Wang W, Wang L, et al. MICRO RNA -21 induces 5-fluorouracil resistance in human pancreatic cancer cells by regulating PTEN and PDCD4. Cancer Med. 2016;5(4):693–702.

[39] Mees ST, Mardin WA, Wendel C, et al. EP300-A miRNA-regulated metastasis suppressor gene in ductal adenocarcinomas of the pancreas: role of EP300 in Metastasis of Pancreatic Cancer. Int J Cancer. 2010;126:114–124.

[40] Kawaguchi T, Komatsu S, Ichikawa D, et al. Clinical impact of circulating miR-221 in plasma of patients with pancreatic cancer. Br J Cancer. 2013;108(2):361–369.

[41] Morimura R, Komatsu S, Ichikawa D, et al. Novel diagnostic value of circulating miR-18a in plasma of patients with pancreatic cancer. Br J Cancer. 2011;105 (11):1733–1740.

[42] Frampton AE, Castellano L, Colombo T, et al. Integrated molecular analysis to investigate the role of microRNAs in pancreatic tumour growth and progression. Lancet. 2015;385:S37.

[43] Giovannetti E, Funel N, Peters GJ, et al. MicroRNA-21 in pancreatic cancer: correlation with clinical outcome and pharmacologic aspects underlying its role in the modulation of gemcitabine activity. Cancer Res. 2010;70(11):4528–4538. .

[44] Frampton AE, Krell J, Jamieson NB, et al. microRNAs with prognostic significance in pancreatic ductal adenocarcinoma: a meta-analysis. Eur J Cancer. 2015;51 (11):1389–1404.

[45] Hwang J-H, Voortman J, Giovannetti E, et al. Identification of microRNA-21 as a biomarker for chemoresistance and clinical outcome following adjuvant therapy in resectable pancreatic cancer. PloS One. 2010;5(5):e10630–e10630. .

[46] Tys LL, Meijer LL, Prado MM, et al. Circulating microRNAs as diagnostic biomarkers for pancreatic cancer. Expert Rev Mol Diagn. 2015;15 (12):1525–1529.

[47] Meijer LL, Garajova I, Caparello C, et al. Plasma miR-181a-5p Downregulation Predicts Response and Improved Survival After FOLFIRINOX in Pancreatic Ductal Adenocarcinoma. Annals of surgery; 2018.

[48] Park J-K, Lee EJ, Esau C, et al. Antisense inhibition of microRNA-21 or −221 arrests cell cycle, induces apoptosis, and sensitizes the effects of gemcitabine in pancreatic adenocarcinoma. Pancreas. 2009;38(7): e190–9.

[49] Toste PA, Li L, Kadera BE, et al. p85α is a microRNA target and affects chemosensitivity in pancreatic cancer. J Surg Res. 2015;196(2):285–293.
[50] Zhao Q, Chen S, Zhu Z, et al. miR-21 promotes EGF-induced pancreatic cancer cell proliferation by targeting Spry2. Cell Death Dis. 2018;9(12):1157.
[51] Zhao M, Wang L, Liu J, et al. MiR-21 Suppresses Anoikis through Targeting PDCD4 and PTEN in Human Esophageal Adenocarcinoma. Curr Med Sci. 2018;38(2):245–251.
[52] Garajová I, Le Large TY, Frampton AE, et al. Molecular mechanisms underlying the role of microRNAs in the chemoresistance of pancreatic cancer. Biomed Res Int. 2014;2014:678401.
[53] Chen S, Chen X, Shan T, et al. MiR-21-mediated Metabolic Alteration of Cancer-associated Fibroblasts and Its Effect on Pancreatic Cancer Cell Behavior. Int J Biol Sci. 2018;14(1):100–110.
[54] Kunita A, Morita S, Irisa TU, et al. MicroRNA-21 in cancer-associated fibroblasts supports lung adenocarcinoma progression. Sci Rep. 2018;8(1):8838.
[55] Czochor JR, Sulkowski P, Glazer PM. miR-155 Overexpression Promotes Genomic Instability by Reducing High-fidelity Polymerase Delta Expression and Activating Error-Prone DSB Repair. Mol Cancer Res. 2016;14(4):363–373.
[56] Ye J, Guo R, Shi Y, et al. miR-155 Regulated Inflammation Response by the SOCS1-STAT3-PDCD4 Axis in Atherogenesis. Mediators Inflamm. 2016;2016:8060182.
[57] Ali S, Banerjee S, Logna F, et al. Inactivation of Ink4a/Arf leads to deregulated expression of miRNAs in K-Ras transgenic mouse model of pancreatic cancer. J Cell Physiol. 2012;227(10):3373–3380.
[58] Zhang J, Manley JL. Misregulation of pre-mRNA alternative splicing in cancer. Cancer Discov. 2013;3(11):1228–1237.
[59] Urbanski LM, Leclair N, Anczuków O. Alternative-splicing defects in cancer: splicing regulators and their downstream targets, guiding the way to novel cancer therapeutics. Wiley Interdiscip Rev RNA. 2018;9:e1476–e1476.
[60] Sciarrillo R, Wojtuszkiewicz A, Assaraf YG, et al. The role of alternative splicing in cancer: from oncogenesis to drug resistance. Drug Resist Updat. 2020;53:100728.
[61] The Human Protein Atlas. SF3B1: pathology Atlas Pancreatic Cancer [Internet]. 2019 [cited 2019 Feb 22]; Available from: https://www.proteinatlas.org/ENSG00000115524-SF3B1/pathology/tissue/pancreatic+cancer#Quantity
[62] The Human Protein Atlas. HNRNPK expression in Pancreatic Cancer [Internet]. 2019; Available from: https://www.proteinatlas.org/ENSG00000165119-HNRNPK/pathology/pancreatic+cancer
[63] He D, Huang C, Zhou Q, et al. HnRNPK/miR-223/FBXW7 feedback cascade promotes pancreatic cancer cell growth and invasion. Oncotarget. 2017;8(12):20165–20178.
[64] Zhou R, Shanas R, Nelson MA, et al. Increased expression of the heterogeneous nuclear ribonucleoprotein K in pancreatic cancer and its association with the mutant p53. Int J Cancer. 2010;126(2):395–404.
[65] Biankin AV, Waddell N, Kassahn KS, et al. Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. Nature. 2012;491(7424):399–405.
[66] Lee SC-W, North K, Kim E, et al. Synthetic Lethal and Convergent Biological Effects of Cancer-Associated Spliceosomal Gene Mutations. Cancer Cell. 2018;34(-225–241.e8):225–241.e8.
[67] Chakedis J, French R, Babicky M, et al. A novel protein isoform of the RON tyrosine kinase receptor transforms human pancreatic duct epithelial cells. Oncogene. 2016;35(25):3249–3259.
[68] Thomas RM, Toney K, Fenoglio-Preiser C, et al. The RON Receptor Tyrosine Kinase Mediates Oncogenic Phenotypes in Pancreatic Cancer Cells and Is Increasingly Expressed during Pancreatic Cancer Progression. Cancer Res. 2007;67(13):6075 LP – 6082.
[69] Christofk HR, Vander Heiden MG, Harris MH, et al. The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. Nature. 2008;452(7184):230–233.
[70] Calabretta S, Bielli P, Passacantilli I, et al. Modulation of PKM alternative splicing by PTBP1 promotes gemcitabine resistance in pancreatic cancer cells. Oncogene. 2016;35(16):2031–2039.
[71] Karni R, De Stanchina E, Lowe SW, et al. The gene encoding the splicing factor SF2/ASF is a proto-oncogene. Nat Struct Mol Biol. 2007;14(3):185–193.
[72] Jensen MA, Wilkinson JE, Krainer AR. Splicing factor SRSF6 promotes hyperplasia of sensitized skin. Nat Struct Mol Biol. 2014;21(2):189–197.
[73] Cohen-Eliav M, Golan-Gerstl R, Siegfried Z, et al. The splicing factor SRSF6 is amplified and is an oncoprotein in lung and colon cancers. J Pathol. 2013;229(4):630–639.
[74] The Human Protein Atlas. SRSF6: Pancreatic cancer pathology [Internet]. 2019; Available from: https://www.proteinatlas.org/ENSG00000124193-SRSF6/pathology/pancreatic+cancer
[75] Li M, Wu P, Yang Z, et al. miR-193a-5p Promotes Pancreatic Cancer Cell Migration and Invasion Through SRSF6-Mediated Alternative Splicing of OGDHL and ECM1. Lancet. 2019;10(1):38–59.
[76] Gonçalves V, Jordan P. Posttranscriptional Regulation of Splicing Factor SRSF1 and Its Role in Cancer Cell Biology. Biomed Res Int. 2015;2015:287048.
[77] Anczuków O, Rosenberg AZ, Akerman M, et al. The splicing factor SRSF1 regulates apoptosis and proliferation to promote mammary epithelial cell transformation. Nat Struct Mol Biol. 2012;19(2):220–228.

[78] Gautrey HL, Tyson-Capper AJ. Regulation of Mcl-1 by SRSF1 and SRSF5 in cancer cells. PloS One. 2012;7 (12):e51497–e51497.

[79] Leu S, Lin Y-M, Wu C-H OP. Loss of Pnn expression results in mouse early embryonic lethality and cellular apoptosis through SRSF1-mediated alternative expression of Bcl-xS and ICAD. J Cell Sci. 2012;125(13):3164 LP – 3172.

[80] Das S, Anczuków O, Akerman M, et al. Oncogenic splicing factor SRSF1 is a critical transcriptional target of MYC. Cell Rep. 2012;1(2):110–117.

[81] Wu H, Sun S, Tu K, et al. A splicing-independent function of SF2/ASF in microRNA processing. Mol Cell. 2010;38(1):67–77.

[82] Woods K, Thomson JM, Hammond SM. Direct Regulation of an Oncogenic Micro-RNA Cluster by E2F Transcription Factors. J Biol Chem. 2007;282 (4):2130–2134.

[83] Xu Q, Li P, Chen X, et al. miR-221/222 induces pancreatic cancer progression through the regulation of matrix metalloproteinases. Oncotarget. 2015;6 (16):14153–14164.

[84] Olive V, Jiang I, He L. mir-17-92, a cluster of miRNAs in the midst of the cancer network. Int J Biochem Cell Biol. 2010;42(8):1348–1354.

[85] Murphy BL, Obad S, Bihannic L, et al. Silencing of the miR-17 92 Cluster Family Inhibits Medulloblastoma Progression. Cancer Res. 2013;73(23):7068–7078.

[86] Olive V, Bennett MJ, Walker JC, et al. miR-19 is a key oncogenic component of mir-17-92. Genes Dev. 2009;23(24):2839–2849.

[87] Kort EJ, Farber L, Tretiakova M, et al. The E2F3-Oncomir-1 axis is activated in Wilms' tumor. Cancer Res. 2008;68(11):4034–4038.

[88] Quattrochi B, Gulvady A, Driscoll DR, et al. MicroRNAs of the mir-17~92 cluster regulate multiple aspects of pancreatic tumor development and progression. Oncotarget. 2017;8(22):35902–35918.

[89] Lin J-C. Therapeutic Applications of Targeted Alternative Splicing to Cancer Treatment. Int J Mol Sci. 2017;19(1):75.

[90] Aslan D, Garde C, Nygaard MK, et al. Tumor suppressor microRNAs are downregulated in myelodysplastic syndrome with spliceosome mutations. Oncotarget. 2016;7(9):9951–9963. .

[91] Pianigiani G, Licastro D, Fortugno P, et al. Microprocessor-dependent processing of splice site overlapping microRNA exons does not result in changes in alternative splicing. RNA. 2018;24(9):1158–1171.

[92] Jang J-Y, Lee Y-S, Jeon Y-K, et al. ANT2 suppression by shRNA restores miR-636 expression, thereby downregulating Ras and inhibiting tumorigenesis of hepatocellular carcinoma. Exp Mol Med. 2013;45(1):e3–e3.

[93] Wang Z, Tong D, Han C, et al. Blockade of miR-3614 maturation by IGF2BP3 increases TRIM25 expression and promotes breast cancer cell proliferation. EBioMedicine. 2019;41:357–369.

[94] Yang R-M, Zhan M, Xu S-W, et al. miR-3656 expression enhances the chemosensitivity of pancreatic cancer to gemcitabine through modulation of the RHOF/EMT axis. Cell Death Dis. 2017;8(10): e3129–e3129.

[95] Itami-Matsumoto S, Hayakawa M, Uchida-Kobayashi S, et al. Circulating Exosomal miRNA Profiles Predict 1570 the Occurrence and Recurrence of Hepatocellular Carcinoma in Patients with Direct-Acting Antiviral-Induced Sustained Viral Response. Biomedicines. 2019;7(4):87

[96] Danan-Gotthold M, Golan-Gerstl R, Eisenberg E, et al. Identification of recurrent regulated alternative splicing events across human solid tumors. Nucleic Acids Res. 2015;43(10):5130–5144.

[97] Kuroyanagi H. Fox-1 family of RNA-binding proteins. Cell Mol Life Sci. 2009;66(24):3895–3907.

[98] Braeutigam C, Rago L, Rolke A, et al. RNA-binding protein Rbfox2: an essential regulator of EMT-driven alternative splicing and a mediator of cellular invasion. Oncogene. 2014;33(9):1082–1092.

[99] Janakiraman H, House RP, Gangaraju VK, et al. (lncRNA) and Short (miRNA) of It: tGFβ-Mediated Control of RNA-Binding Proteins and Noncoding RNAs. Mol Cancer Res. 2018;16(4):567 LP – 579.

[100] Zhang Y, Li M, Wang H, et al. Profiling of 95 MicroRNAs in Pancreatic Cancer Cell Lines and Surgical Specimens by Real-Time PCR Analysis. World J Surg. 2009;33(4):698–709.

[101] Frampton AE, Castellano L, Colombo T, et al. MicroRNAs Cooperatively Inhibit a Network of Tumor Suppressor Genes to Promote Pancreatic Tumor Growth and Progression. Gastroenterology. 2014;146(1):268–277.e18.

[102] Bloomston M, Frankel WL, Petrocca F, et al. MicroRNA Expression Patterns to Differentiate Pancreatic Adenocarcinoma From Normal Pancreas and Chronic Pancreatitis. JAMA. 2007;297(17):1901.

[103] Wang YJ, McAllister F, Bailey JM, et al. Is Required for Maintenance of Adult Pancreatic Acinar Cell Identity and Plays a Role in Kras-Driven Pancreatic Neoplasia. PLoS ONE. 2014;9(11):e113127.

[104] Morris JP, Greer R, Russ HA, et al. Dicer Regulates Differentiation and Viability during Mouse Pancreatic Cancer Initiation. PLoS ONE. 2014;9(5):e95486.

[105] Rachagani S, Macha MA, Menning MS, et al. Changes in microRNA (miRNA) expression during pancreatic cancer development and progression in a genetically engineered KrasG12D;Pdx1-Cre mouse (KC) model. Oncotarget. 2015;6(37):40295–40309.

[106] Kędzierska H, Piekiełko-Witkowska A. Splicing factors of SR and hnRNP families as regulators of apoptosis in cancer. Cancer Lett. 2017;396:53–65.

[107] Gonçalves V, Pereira JFS, Jordan P. Signaling Pathways Driving Aberrant Splicing in Cancer Cells. Genes (Basel). 2017;9(1):9.

[108] Qiao L, Xie N, Bai Y, et al. Identification of Upregulated HNRNPs Associated with Poor Prognosis in Pancreatic Cancer. Biomed Res Int. 2019;2019:5134050.
[109] Chen Z-Y, Cai L, Zhu J, et al. Fyn requires HnRNPA2B1 and Sam68 to synergistically regulate apoptosis in pancreatic cancer. Carcinogenesis. 2011;32(10):1419–1426.
[110] Shen K, Cao Z, Zhu R, et al. The dual functional role of MicroRNA-18a (miR-18a) in cancer development. Clin Transl Med. 2019;8(1):32.
[111] Tsang WP, Kwok TT. The miR-18a* microRNA functions as a potential tumor suppressor by targeting on K-Ras. Carcinogenesis. 2009;30(6):953–959.
[112] Cogoi S, Rapozzi V, Cauci S, et al. Critical role of hnRNP A1 in activating KRAS transcription in pancreatic cancer cells: a molecular mechanism involving G4 DNA. Biochim Biophys Acta Gen Subj. 2017;1861(5):1389–1398.
[113] Mikula M, Dzwonek A, Karczmarski J, et al. Landscape of the hnRNP K protein–protein interactome. PROTEOMICS. 2006;6(8):2395–2406.
[114] Ma J, Cao T, Cui Y, et al. miR-223 Regulates Cell Proliferation and Invasion via Targeting PDS5B in Pancreatic Cancer Cells. Mol Ther Nucleic Acids. 2019;14:583–592.
[115] Ma J, Cheng L, Liu H, et al. Genistein Down-Regulates miR-223 Expression in Pancreatic Cancer Cells. Curr Drug Targets. 2013;14(10):1150–1156.
[116] Li C, Zhao Z, Zhou Z, et al. Linc-ROR confers gemcitabine resistance to pancreatic cancer cells via inducing autophagy and modulating the miR-124/PTBP1/PKM2 axis. Cancer Chemother Pharmacol. 2016;78(6):1199–1207.
[117] Wu D-H, Liang H, Lu S-N, et al. miR-124 Suppresses Pancreatic Ductal Adenocarcinoma Growth by Regulating Monocarboxylate Transporter 1-Mediated Cancer Lactate Metabolism. Cell Physiol Biochem. 2018;50(3):924–935.
[118] Wang P, Chen L, Zhang J, et al. Methylation-mediated silencing of the miR-124 genes facilitates pancreatic cancer progression and metastasis by targeting Rac1. Oncogene. 2014;33(4):514–524.
[119] Yeom K-H, Mitchell S, Linares AJ, et al. Polypyrimidine tract-binding protein blocks miRNA-124 biogenesis to enforce its neuronal-specific expression in the mouse. PNAS. 2018; 115:E11061–70.
[120] Wu D, Pan H, Zhou Y, et al. microRNA-133b down-regulation and inhibition of cell proliferation, migration and invasion by targeting matrix metallopeptidase-9 in renal cell carcinoma. Mol Med Rep. 2014;9(6):2491–2498.
[121] Wang Q-Y, Zhou C-X, Zhan M-N, et al. MiR-133b targets Sox9 to control pathogenesis and metastasis of breast cancer. Cell Death Dis. 2018;9(7):752.
[122] Li D, Xia L, Chen M, et al. miR-133b, a particular member of myomiRs, coming into playing its unique pathological role in human cancer. Oncotarget. 2017;8 (30):50193–50208.
[123] Khan MA, Zubair H, Srivastava SK, et al. Insights into the Role of microRNAs in Pancreatic Cancer Pathogenesis: potential for Diagnosis, Prognosis, and Therapy. Adv Exp Med Biol. 2015;889:71–87.
[124] Taniguchi K, Sakai M, Sugito N, et al. PTBP1-associated microRNA-1 and −133b suppress the Warburg effect in colorectal tumors. Oncotarget. 2016;7(14):18940–18952.
[125] Siddika T, Heinemann IU. Bringing MicroRNAs to Light: methods for MicroRNA Quantification and Visualization in Live Cells. Front Bioeng Biotechnol. 2021;8:619583.
[126] Dave VP, Ngo TA, Pernestig A-K, et al. MicroRNA amplification and detection technologies: opportunities and challenges for point of care diagnostics. Lab Invest. 2019;99(4):452–469.
[127] Ye J, Xu M, Tian X, et al. Research advances in the detection of miRNA. J Pharm Anal. 2019;9 (4):217–226.
[128] Saliminejad K, Khorram Khorshid HR, Soleymani Fard S, et al. An overview of microRNAs: biology, functions, therapeutics, and analysis methods: SALIMINEJAD ET AL. J Cell Physiol. 2019;234 (5):5451–5465.
[129] Chu Y-H, Hardin H, Zhang R, et al. In situ hybridization: introduction to techniques, applications and pitfalls in the performance and interpretation of assays. Semin Diagn Pathol. 2019; 36:336–341.
[130] Kilic T, Erdem A, Ozsoz M, et al. microRNA biosensors: opportunities and challenges among conventional and commercially available techniques. Biosens Bioelectron. 2018;99:525–546.
[131] Ouyang T, Liu Z, Han Z, et al. MicroRNA Detection Specificity: recent Advances and Future Perspective. Anal Chem. 2019;91(5):3179–3186.
[132] Gines G, Menezes R, Xiao W, et al. Emerging isothermal amplification technologies for microRNA biosensing: applications to liquid biopsies. Mol Aspects Med. 2020;72:100832.
[133] Jet T, Gines G, Rondelez Y, et al. Advances in multiplexed techniques for the detection and quantification of microRNAs. Chem Soc Rev. 2021;10:1039.
[134] Rondelez Y, Gines G. Multiplex Digital MicroRNA Detection Using Cross-Inhibitory DNA Circuits. ACS Sensors. 2020;5(8):2430–2437.
[135] Sciarrillo R, Wojtuszkiewicz A, El Hassouni B, et al. Splicing modulation as novel therapeutic strategy against diffuse malignant peritoneal mesothelioma. EBioMedicine; 2018.
[136] Dehm SM. Test-Firing Ammunition for Spliceosome Inhibition in Cancer. Clin Cancer Res. 2013;19 (22):6064 LP – 6066.
[137] Hong DS, Kurzrock R, Naing A, et al. A phase I, open-label, single-arm, dose-escalation study of E7107, a precursor messenger ribonucleic acid (pre-mRNA) splicesome inhibitor administered intravenously on

days 1 and 8 every 21 days to patients with solid tumors. Invest New Drugs. 2014;32(3):436–444.

[138] Zhang Q, Di C, Yan J, et al. Inhibition of SF3b1 by pladienolide B evokes cycle arrest, apoptosis induction and p73 splicing in human cervical carcinoma cells. Artif Cells Nanomed Biotechnol. 2019;47 (1):1273–1280.

[139] Zhang Y, Yuan Z, Jiang Y, et al. Inhibition of Splicing Factor 3b Subunit 1 (SF3B1) Reduced Cell Proliferation, Induced Apoptosis and Resulted in Cell Cycle Arrest by Regulating Homeobox A10 (HOXA10) Splicing in AGS and MKN28 Human Gastric Cancer Cells. Med Sci Monit. 2020;26:e919460.

[140] Hepburn LA, McHugh A, Fernandes K, et al. Targeting the spliceosome for cutaneous squamous cell carcinoma therapy: a role for c-MYC and wild-type p53 in determining the degree of tumour selectivity. Oncotarget. 2018;9(33):23029–23046.

[141] Eskens FALM, Ramos FJ, Burger H, et al. Phase I Pharmacokinetic and Pharmacodynamic Study of the First-in-Class Spliceosome Inhibitor E7107 in Patients with Advanced Solid Tumors. Clin Cancer Res. 2013;19(22):6296–6304.

[142] Steensma DP, Wermke M, Klimek VM, et al. Results of a Clinical Trial of H3B-8800, a Splicing Modulator, in Patients with Myelodysplastic Syndromes (MDS), Acute Myeloid Leukemia (AML) or Chronic Myelomonocytic Leukemia (CMML). Blood. 2019;134(Supplement_1):673.

[143] Amin E, Oltean S, Hua J, et al. WT1 Mutants Reveal SRPK1 to Be a Downstream Angiogenesis Target by Altering VEGF Splicing. Cancer Cell. 2011;20 (6):768–780.

[144] Hatcher JM, Wu G, Zeng C, et al. SRPKIN-1: a Covalent SRPK1/2 Inhibitor that Potently Converts VEGF from Pro-angiogenic to Anti-angiogenic Isoform. Cell Chem Biol. 2018;25(4):460–470.e6.

[145] Iwatani-Yoshihara M, Ito M, Klein MG, et al. Discovery of Allosteric Inhibitors Targeting the Spliceosomal RNA Helicase Brr2. J Med Chem. 2017;60(13):5759–5771.

[146] Jagtap PKA, Garg D, Kapp TG, et al. Rational Design of Cyclic Peptide Inhibitors of U2AF Homology Motif (UHM) Domains To Modulate Pre-mRNA Splicing. J Med Chem. 2016;59(22):10190–10197.

[147] Effenberger KA, James RC, Urabe VK, et al. The Natural Product N-Palmitoyl-l-leucine Selectively Inhibits Late Assembly of Human Spliceosomes. J Biol Chem. 2015;290(46):27524–27531.

[148] Sidarovich A, Will CL, Anokhina MM, et al. Identification of a small molecule inhibitor that stalls splicing at an early step of spliceosome activation. Elife. 2017;6:e23533.

[149] Kotake Y, Sagane K, Owa T, et al. Splicing factor SF3b as a target of the antitumor natural product pladienolide. Nat Chem Biol. 2007;3(9):570.

[150] Teufel A, Krupp M, Weinmann A, et al. Current bioinformatics tools in genomic biomedical research (Review). International Journal of Molecular Medicine [Internet] 2006 [cited 2020 Oct 20]; Available from: http://www.spandidos-publications.com/10.3892/ijmm.17.6.967

[151] Shen S, Park JW, Lu Z, et al. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. PNAS. 2014;111:E5593–601.

[152] Trincado JL, Entizne JC, Hysenaj G, et al. SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. Genome Biology [Internet] 2018 [cited 2020 Oct 20]; 19. Available from: https://genomebiology.biomedcentral.com/articles/10.1186/s13059-018-1417-1

[153] Katz Y, Wang ET, Airoldi EM, et al. Analysis and design of RNA sequencing experiments for identifying isoform regulation. Nat Methods. 2010;7(12):1009–1015.

[154] Bailey TL, Boden M, Buske FA, et al. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 2009;37(Web Server):W202–8.

[155] Kazan H, Ray D, Chan ET, et al. RNAcontext: a New Method for Learning the Sequence and Structure Binding Preferences of RNA-Binding Proteins. PLoS Comput Biol. 2010;6(7):e1000832.

[156] Huang H-Y, Lin Y-C-D, Li J, et al. miRTarBase 2020: updates to the experimentally validated microRNA–target interaction database. Nucleic Acids Res 2019 gkz896 10.1093/nar/gkz896

[157] Chen Y, Wang X. miRDB: an online database for prediction of functional microRNA targets. Nucleic Acids Res. 2020;48(D1):D127–31.

[158] Lahat A, Grellscheid SN Differential mRNA Alternative Splicing [Internet]. In: Aransay AM, Lavín Trueba JL, editors. Field Guidelines for Genetic Experimental Designs in High-Throughput Sequencing. Cham: Springer International Publishing; 2016 [cited 2020 Oct 20]. page 105–119.Available from: http://link.springer.com/10.1007/978-3-319-31350-4_5

[159] Bodén M, Bailey TL. Associating transcription factor-binding site motifs with target GO terms and target genes. Nucleic Acids Res. 2008;36(12):4108–4117.

[160] Maticzka D, Lange SJ, Costa F, et al. GraphProt: modeling binding preferences of RNA-binding proteins. Genome Biol. 2014;15(1):R17.

[161] Munteanu A, Mukherjee N, Ohler U SSMART: sequence-structure motif identification for RNA-binding proteins. Bioinformatics [Internet] 2018 [cited 2020 Oct 20]; Available from: https://academic.oup.com/bioinformatics/advance-article/doi/10.1093/bioinformatics/bty404/5035778

[162] Dang LT, Tondl M, Chiu MHH, et al. TrawlerWeb: an online de novo motif discovery tool for next-generation sequencing datasets. BMC Genomics [Internet] 2018 [cited 2020 Oct 20]; 19. Available from: https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-018-4630-0

[163] Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences to function. Nucleic Acids Res. 2019;47(D1):D155–62.

[164] Karagkouni D, Paraskevopoulou MD, Chatzopoulos S, et al. DIANA-TarBase v8: a decade-long collection of experimentally supported miRNA–gene interactions. Nucleic Acids Res. 2018;46(D1):D239–45.

[165] Ru Y, Kechris KJ, Tabakoff B, et al. The multiMiR R package and database: integration of microRNA–target interactions along with their disease and drug associations. Nucleic Acids Res. 2014;42(17):e133–e133.

[166] Fromm B, Domanska D, Høye E, et al. MirGeneDB 2.0: the metazoan microRNA complement. Nucleic Acids Res. 2020;48(D1):D132–41.

[167] Hackenberg M, Rodriguez-Ezpeleta N, Aransay AM. miRanalyzer: an update on the detection and analysis of microRNAs in high-throughput sequencing experiments. Nucleic Acids Res. 2011;39(suppl):W132–8.

[168] Friedländer MR, Mackowiak SD, Li N, et al. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic Acids Res. 2012;40(1):37–52.

[169] Aparicio-Puerta E, Lebrón R, Rueda A, et al. sRNAbench and sRNAtoolbox 2019: intuitive fast small RNA profiling and differential expression. Nucleic Acids Res. 2019;47(W1):W530–5.

[170] Langmead B, Trapnell C, Pop M, et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10(3):R25.

[171] Dillhoff M, Liu J, Frankel W, et al. MicroRNA-21 is Overexpressed in Pancreatic Cancer and a Potential Predictor of Survival. J Gastrointestinal Surg. 2008;12 (12):2171–2176.

[172] Cote GA, Gore JA, McElyea SD, et al. Study to Develop a Diagnostic Test for Pancreatic Ductal Adenocarcinoma Based on Differential Expression of Select miRNA in Plasma and Bile. Am J Gastroenterol. 2014;109(12):1942–1952

[173] Michael Traeger M, Rehkaemper J, Ullerich H, et al. The ambiguous role of microRNA-205 and its clinical potential in pancreatic ductal adenocarcinoma. J Cancer Res Clin Oncol. 2018;144(12):2419–2431.

[174] Qin R-F, Zhang J, Huo H-R, et al. MiR-205 mediated APC regulation contributes to pancreatic cancer cell proliferation. World J Gastroenterol. 2019;25(28):3775–3786.

[175] Chang W, Liu M, Xu J, et al. MiR-377 inhibits the proliferation of pancreatic cancer by targeting Pim-3. Tumor Biol. 2016;37(11):14813–14824.

[176] Yu Y, Liu L, Ma R, et al. MicroRNA-127 is aberrantly downregulated and acted as a functional tumor suppressor in human pancreatic cancer. Tumor Biol. 2016;37(10):14249–14257.

[177] Zhang G, Liu D, Long G, et al. Downregulation of microRNA-181d had suppressive effect on pancreatic cancer development through inverse regulation of KNAIN2. Tumor Biol. 2017;39:101042831769836.

[178] Imamura T, Komatsu S, Ichikawa D, et al. Depleted tumor suppressor miR-107 in plasma relates to tumor progression and is a novel therapeutic target in pancreatic cancer. Scientific Reports [Internet] 2017 [cited 2020 Oct 29]; 7. Available from: http://www.nature.com/articles/s41598-017-06137-8

[179] Tavano F, Gioffreda D, Valvano MR, et al. Droplet digital PCR quantification of miR-1290 as a circulating biomarker for pancreatic cancer. Scientific Reports [Internet] 2018 [cited 2020 Oct 29]; 8. Available from: http://www.nature.com/articles/s41598-018-34597-z

[180] Chen Y, Yang F, Zubovic L, et al. Targeted inhibition of oncogenic miR-21 maturation with designed RNA-binding proteins. Nat Chem Biol. 2016;12(9):717–723.

[181] Chen Y, Zubovic L, Yang F, et al. Rbfox proteins regulate microRNA biogenesis by sequence-specific binding to their precursors and target downstream Dicer. Nucleic Acids Res. 2016;44(9):4381–4395.

Supplementary Appendix Chapter 10

Interrelationship between miRNA and splicing factors in pancreatic ductal adenocarcinoma

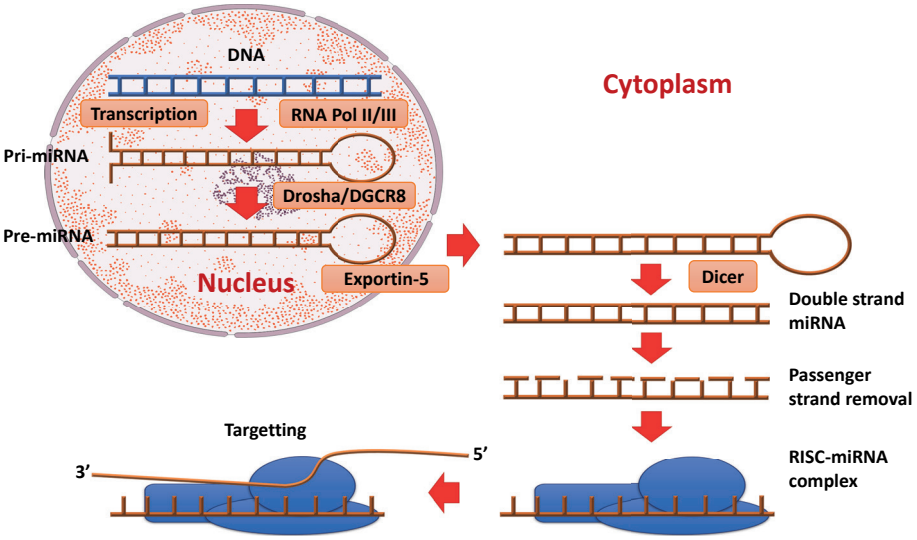


Figure S1. The biosynthesis process of miRNA. miRNA is transcribed from DNA by RNA pol II which produced a hairpin-shaped transcript. DROSHA removes the overhanging sequence and exportin-5 transports the pre-miRNA to cytoplasm where Dicer removes the loop section. After removal of passenger strand, mature miRNA forms the RISC complex with Ago proteins which can target and degrade any mRNA with compatible sequence[21,22].

Chapter 11

Exploring splicing modulation as a novel strategy against pancreatic cancer

Mantini G*, Sciarrillo R*, Randazzo O*, El Hassouni B, Supadmanaba IGP, Lagerweij T, Würdinger T, Peters GJ, Molthoff CFM, Jansen G, Kaspers GJL, Cloos J, Giovannetti E
*These authors contributed equally

Manuscript in preparation

Chapter 12

DISCUSSION AND FUTURE PERSPECTIVES

Cancer is a complex and heterogeneous disease. Many different factors such as genetic aberrations, epigenetic alterations, hijacked cellular signaling, metabolic alterations, and environmental factors can influence the onset of tumors. Thus, there is a need for large-scale data of each tumor type that includes different tumor stages and treatments at multiple data layers (epigenomics, genomics, proteomics, metabolomics, etc.) and from diverse populations.

Several consortia are already taking action to provide an atlas of cancer data to the research community[1–4]. Despite the increasing number of molecular data for multiple tumor types and matched normal tissues at different -omics levels, the downstream analysis is often challenging and left up to the researcher.

The difficulties of data integration certainly include the type of -omics data, number of samples, infrastructures, availability of standardized pipelines, technical problems (low quality and error rates), and type of validation. However, successful integration of -omics data paves for a better understanding of this complex disease.

In recent years, the focus on precision medicine has shifted from genomic profiling to a more comprehensive assessment of tumor biology by integrating genomic, transcriptomic, and proteomic data[5–7]. Multi-omics analysis has yielded detailed insights into the heterogeneity and complexity of cancers. The main strength of -omics data integration is the capacity to track the flow from upstream information (epigenomics, genomics) to downstream effects (proteomics, phosphoproteomics, metabolomics).

This thesis aimed to explore Pancreatic Ductal Adenocarcinoma (PDAC) from different -omics views of various clinically relevant sample types (tumor tissues, microdissected cells, blood platelets) to understand its aggressive biology and identify novel biomarkers for disease diagnosis, prognosis, or to guide specific therapies.

The first part of this thesis reported novel markers for PDAC prognosis from tissue proteomics analyses and should pave the way to new integrative strategies with phosphoproteomics and other -omics data.

In particular, we employed co-expression analysis (WGCNA) and LCM coupled to mass-spectrometry to identify prognostic markers from bulk tumors and stromal and epithelial components, specifically.

The second part of this thesis evaluated the use of multiple types of RNAs as drug targets for aberrant PDAC metabolism or candidate biomarkers for distinguishing PDAC from other benign and malignant diseases and predict PDAC stages using blood platelets. In particular, the biology of PDAC platelets is explored through multiple -omics levels to better understand the mechanism of action of tumor-educated platelets.

Finally, the third part of this thesis discussed strategies for drug combination and selection of specific therapy through biomarker expression. Moreover, the use of novel splicing inhibitors is extensively evaluated for future application in the clinic for PDAC treatment.

Proteomics and their application in cancer research and PDAC specifically

The development of cancer is typically associated with genetic and epigenetic alterations. However, from a functional perspective, cancer is a disease of proteins where the phenotype is often the result of deranged signaling pathways[8].

Thus, proteomics and phosphoproteomics studies are essential tools that can be analyzed in an integrative manner with genomic studies to decipher this deadly and complex disease. In **Chapter 2**, several computational methods for multi-omics integration with phosphoproteomics data are reviewed. In addition, we included the state-of-the-art methods for kinase activity analysis, the latest large-scale proteogenomics studies with phosphoproteomics data, and bioinformatics methods where phosphoproteomics can be used as another -omics layer. We also expanded on a successful multi-omics data integration where phosphoproteomics was the key layer to distinguish two early-onset gastric cancer subtypes and elucidate drug sensitivity prediction. In particular, this review provided readers with existing algorithms and tools for multi-omics data integration to be used for phosphoproteomics data integration. We expect that research in this direction will benefit from advances in mass spectrometry-based methods for more comprehensive site-specific phosphorylation profiling and the ability to integrate all available phosphorylation datasets for the analysis of individual patients at hand. However, developing new tools for -omics data integration remains a constantly evolving field.

In **Chapter 3,** a co-expression network analysis method was applied to the proteome data of resected PDAC patients to infer biological functions and novel prognostic PDAC biomarkers. Protein co-expression modules were linked to well-known PDAC hallmarks of cancer such as axon-guidance, EMT, oxidative phosphorylation, MYC targets, and KRAS signaling. Notably, one module was found to be significantly associated with survival. This module was functionally enriched for glycolysis, EMT, apoptosis, and reactive oxidative stress, highlighting a possible interplay between these biological processes. Three prognostic markers were evaluated and validate from this module: SPTBN1, KHSRP, and PYGL. These markers may be used in the future to predict the clinical outcome of PDAC resected patients.

Moreover, four modules showed overrepresentation of different sets of EMT genes associated with metabolic pathways, suggesting that cell metabolism can influence the EMT state or vice versa. Of note, previous studies already showed the association of tumor metabolism and EMT[9,10], illustrating the complex interplay between EMT and metabolic reprogramming. Interestingly, these four modules were regulated by different transcription factors. In the magenta module, transcription factors binding sites (TFBS) for STAT5 and E12 were noted. STAT5 is overexpressed during EMT, and aberrant activity of this transcription factor has been found to induce mitochondrial dysfunction and reactive oxygen species (ROS) formation, leading to DNA damage[11]. Besides, E12 is associated with the repression of E-cadherin (and thus EMT) in mouse models[12]. Although we captured three new prognostic biomarkers and the biology related to these, some limitations to our study need to be noted. A limit is certainly the small sample size (n=20) and the analysis of PDAC tumor as a whole. Separation of the epithelial tissue and stromal part of bulk tumors can be done by Laser capture Microdissection

techniques (LCM) or using recent tumor deconvolution tools. The latter is a novel research field that has recently emerged to avoid the expensive costs of single-cell technologies and use existing public data based on bulk samples. Although this type of research is increasingly applied to genomics data, there are no studies on tumor deconvolution in proteomics.

Chapter 4 describes the use of LCM on PDAC tumors to isolate the epithelial and stromal compartments for further proteomics analysis. Until recently, the combination of low sample input with mass spectrometry posed a challenge, as amplification of proteins is not possible. However, the sensitivity of mass spectrometers has improved over the last decade, creating options for coupling LCM of tumors to proteomics. With this approach, two stromal subtypes were identified from the stromal compartment and aligned with the existing stromal subtypes based on transcriptomics data. COL11A1 was identified as a poor prognostic stromal marker, and its high expression has been determined in multiple gene expression sets correlated to poor prognostic cancer types.

On the other hand, proteomics data did not reveal the epithelial subtypes previously identified using transcriptomics data. This result suggested an alteration of information flow from gene transcripts to protein expression and was assessed by analyzing gene-protein correlations of PDAC mouse models. Importantly, CALB2 was identified and validated as a prognostic marker in the tumor compartment. Of note, this marker was already shown to be prognostic in mesothelioma[13]. Furthermore, by enriching for tumor areas (epithelial compartment), the role of EPHA2 was uncovered and functionally validated as a therapeutic target in PDAC. Interestingly, EPHA2 was recently identified as a key regulator of the PDAC immune-suppressive microenvironment[14]. Of note, this study has resulted in the deepest PDAC proteome to date and will be a relevant source for future studies based on tumor deconvolution from proteomics data.

KEY POINTS

- The phosphoproteome provides a unique layer of information that can help in subtype discovery and drug target identification.
- Correlation networks can extrapolate clinically relevant information taking into account the associated biology of multiple proteins.
- LCM and tumor deconvolution studies may help in disentangle PDAC complexity and tumor-stroma crosstalk.

Studies on multiple types of RNA and Platelets

The poor prognosis of PDAC patients is due to the inherent and/or acquired resistance to conventional treatment modalities [15]. Despite the constant efforts to formulate new chemotherapy regimens, novel strategies for personalized treatments are urgently warranted[16,17]. In **Chapter 5,** we summarize the main therapeutic options for PDAC and then critically discuss the use of microRNAs as novel potential biomarkers to predict aberrant metabolism and optimize new therapeutic strategies. In conclusion, future studies should evaluate the role of several emerging miRNAs as biomarkers in PDAC and validate their use as

potential predictive factors for the response to glycolysis inhibitors. Moreover, additional candidate miRNAs could be found by investigating the metabolism of glycolysis inhibitors. Indeed, miRNAs could modulate the expression of several drug-metabolizing enzymes, affecting the sensitivity to this class of new drugs. Hopefully, identifying biomarkers of response, including the miRNAs mentioned above, should shorten the gap between preclinical studies and personalized therapies using these novel treatments.

An emerging field in PDAC prognosis and diagnosis is the use of minimally invasive tools. Currently, diverse liquid biopsy sources are under investigation for PDAC diagnosis, such as circulating tumor DNA (ctDNA), cell-free DNA methylation (cfDNA), miRNAs, protein markers, and intron-spanning RNA reads from blood platelets. [18–23] In this regard, **Chapter 6** demonstrates how the analysis of intron-spanning reads from tumor-educated platelets (TEPs) can help distinguish PDAC patients from benign disease. Moreover, this model demonstrated to classify PDAC patients to early or late stage, where pre-clinically discovered biomarkers were often insufficiently sensitive to detect early-stage tumors. Although only clinically relevant diagnostic groups were included, the diagnostic classifier can be optimized by adding more features (e.g., each benign disease group). Additionally, large-scale validation is already ongoing.

Tumor diagnosis from a single blood withdrawal holds great potential in the clinic, and the recent results from blood platelet RNA classifiers are very promising[23–25]. Despite those positive results, the biology of tumor-educated platelets is still unknown. **Chapter 7** describes an integrative analysis of miRNAs, transcriptomics, and proteomics data of PDAC and benign disease blood platelets. This study aimed to understand how platelets can change their expression profile when "educated" by the tumor. Moreover, this study shows that platelets change their expression profile in patients with PDAC through dysregulation of miRNAs and splicing factors, supporting the presence of *de novo* protein machinery that can "educate" the platelets. Gene ontology mining revealed enrichment of RNA splicing, mRNA processing, and translation initiation in miRNAs and proteins of PDAC educated platelets. Remarkably, modified forms of miRNAs involved in cancer signaling are found to act on the down-regulation of SPARC, a tumor suppressor gene. These novel findings could indeed be further exploited for innovative liquid biopsies platforms as well as possible therapeutic targets. Overall, this was the first study where miRNAs, RNAs, and proteins were profiled from the same set of platelet samples in PDAC patients. Considering the sample size limitation, this study aimed to explore platelets' intrinsic regulation using -omics profiling. To this end, an extensive catalog of molecular profiles and an interactive tool to visualize expected correlations were generated to facilitate further investigations on additional diagnostic biomarkers and therapeutic tools.

KEY POINTS

- Enhanced splicing activity in PDAC platelets is shown at different -omics levels.
- miR-29a-3p and its isomiR forms are involved in the down-regulation of the tumor suppressor SPARC in PDAC blood platelets.

Pharmacology in PDAC

Combination therapies are used to achieve a cure, better efficacy, and circumvent resistant disease in patients. **Chapter 8** summarizes the types of analyses that can be used to determine the efficacy of drug combinations prior test. In general, various *in vitro* models and mathematical equations based on the Bliss independence model, Loewe additivity model, or a combination of both, *i.e.,* the median drug effect, enable predicting the effect of two agents combined. Nonetheless, applying these models should be tailored to the context of each study and with an awareness of the limitations and advantages of each method. Considering the increment in the development of novel therapeutics, the number of combinations that can be tested is substantial. More advanced computational models such as deep-learning methodologies might be needed. Therefore, refinements of the classical synergy models with emerging artificial intelligence-based algorithms will benefit the investigation of new combination treatments in the near future.

While novel drug combinations are under study, the current regimen for PDAC treatment uses gemcitabine (alone or in combination with nab-paclitaxel) or FOLFIRINOX. The latter regimen is increasingly being used in both palliative and adjuvant settings but cannot be used for all PDAC patients. For this reason, the discovery of new biomarkers for patient stratification may indeed pave the road to the selection of patients who are likely to receive a benefit from gemcitabine-based adjuvant chemotherapy, and therefore has immediate clinical relevance. **Chapter 9** reviews the predictive value of hENT-1 for gemcitabine activity in PDAC and its contradictory data. The reliability of a predictive biomarker is assessed through the analysis of sensitivity, specificity, and accuracy values. However, these parameters need extensive validation studies and quality assessment before approval and application in a clinical setting. Several studies support the association between no or low hENT-1 expression in tumors and poor response to gemcitabine in other cancers, including bladder, biliary tract, and lung cancers. Thus, the validation of hENT-1 in standardized IHC techniques would be helpful also for other cancer types. However, identifying solid and reproducible molecular markers is amongst the biggest challenges in personalized cancer medicine. Therefore, in **Chapter 9**, molecular mechanisms influencing the expression and activity of hENT-1 were also reported.

Next to emerging biomarkers for patient therapy stratification, splicing inhibitors are under investigation as a novel therapeutic strategy. **Chapter 10** extensively reviewed the interaction between splicing and miRNA, as well as new bioinformatics tools to predict the effect of splicing modulation toward miRNA profiles. For instance, the upregulation of tumor-suppressing miRNAs may mediate an anticancer effect of splicing modulation, as was shown by the inhibition of PTBP1 and SF3B1 in cervical cancer. Potentially similar inhibitory effects and the impact of other SFs on cancer progression and miRNA profile still need to be investigated in PDAC. Moreover, specific miRNAs could be used as a target to downregulate specific SFs and also for combined therapeutic approaches.

Novel bioinformatics tools provide extensive data that can be used to deepen our knowledge in the biological effects of the interplay between miRNAs and splicing mechanisms, as well as several predictive models for target prediction to prioritize future experimental and clinical

validation. Therefore, targeting aberrant splicing and the reciprocal interaction with deregulated miRNA could consequently provide more effective therapeutic approaches to combat the complex biology of PDAC and its chemoresistance features.
In **Chapter 11**, we evaluated the effect of splicing modulator E7107 in orthotopic cancer mouse models. Overexpression of the splicing factor SF3B1 is often seen in pancreatic cancer tissues, and its modulation has never been evaluated in this tumor. SF3B1 resulted in being a potential prognostic factor as well as a therapeutic target for PDAC, where its modulation affects proliferation and migration.

KEY POINTS

- Emerging artificial intelligence-based models will benefit the identification of new combination treatments in the near future
- The splicing modulator E7107 showed anti-tumor activity in orthotopic cancer mouse models supporting further studies and novel therapeutic opportunities in PDAC with splicing inhibitors

Future perspective
In this thesis, we presented a number of markers for PDAC prognosis and drug resistance, as well as a panel of spliced-RNAs for PDAC classification from benign disease. The tumor biology of PDAC was also explored through a multiple -omics view: proteomics from tumor tissue (bulk, stroma, and epithelial compartment), genomics (transcriptomics and miRNAs), and proteomics from blood platelets of PDAC and benign patients.
As demonstrated in this thesis, each biomarker might be a potential candidate for prognosis and possibly, decision treatment.
Interestingly, an expression panel of intron-spanning RNA reads of blood platelets distinguished PDAC patients from benign lesions. The results from this study hold promising opportunities for the employment of Artificial Intelligence (AI) in clinical practice. Moreover, AI can circumvent the problem of analyzing multiple molecules simultaneously, such as the integration of cfDNA, RNA, and proteins. This “multi-omics panel” might be a more robust marker than single markers that are often used nowadays. Unfortunately, we are far away from the actual implementation of AI in clinical practice. The main question is: “How do you know the machine is right?”. At the moment, AI can help direct doctors towards the right therapy or decision to take for the patient.
However, most of the studies presented in this thesis shared a common weakness: the limited number of samples. The small sample size is certainly a statistical limitation. Still, it has to be considered the very small fraction of patients diagnosed with PDAC at early or resectable stage (20%), which was the type of patient mainly studied in this thesis.
It is well known that PDAC is often diagnosed at a very late stage, and this has to be taken into account in future studies design. Hence, we need data depth in the number of samples to

optimize PDAC cancer research and personalized treatments [26,27, 33]. In particular, multi-center data integration can circumvent this problem.

Data gathering from multi-center studies has already been initiated from the National Cancer Institute (NIH) in 2005 when announcing the "Human Cancer Genome Project" to obtain a comprehensive understanding of the genomic alterations that underlie all major cancers. Today, this is known as the Cancer Research Data Commons Repositories (CRDC) [28]. These repositories contain data that have been harmonized and stored in a format ready for analysis and offer a consolidated network for accessing data, tools, and workflows across scientific domains. Currently, there are three accessible repositories: (i) Genomic Data Commons (GDC) that includes cancer genomic profiling (TCGA), its pediatric equivalent (TARGET), the Cancer Cell Line Encyclopedia (CCLE), and continuously other growing sources; (ii) Proteomic Data Commons that includes protein profile of clinical tumor samples (CPTAC); (iii) Imaging Data Commons (IDC) which provides a wide variety of medical imaging and metadata from NCI projects.

Providing data repositories with a consistent number of samples is the first step towards data integration. Recently, an increasing trend of integrative studies has been reported, especially in the oncology field. However, data integration requires expertise in data analysis and computational biology, as well as specific knowledge in the lab protocols adopted and the tumor type under consideration.

From the computational side, EBI (European Bioinformatics Institute) and ELIXIR are joining efforts to help scientists across Europe to safely store and share their data, supporting existing databases or computing facilities in individual countries[29,30]. In particular, ELIXIR host a comprehensive registry of software, databases, and tools to facilitate scientists worldwide to find, understand, utilize the resources and tools they need.

As technology continues to progress, we might soon be able to obtain a vast amount of data for each cancer patient: from the treatment course, recurrence, genomic composition, disease history. This massive data on a large patient population will dissect common malignancies into distinct subtypes and deliver precision oncology treatment with higher accuracy and efficiency[31] (**Figure 1**).

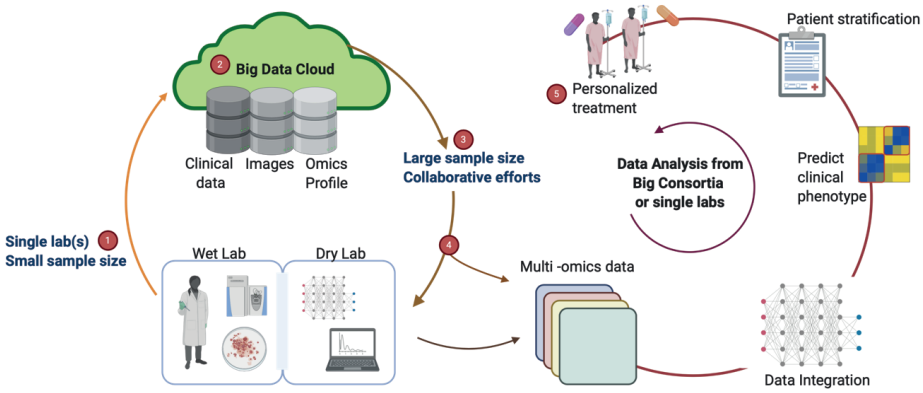


Figure 1. Schematic representation of oncology research when multiple labs collaborate to boost quantity and quality of research. Worldwide, labs usually produce small and typically single layer -omics data (wet lab). The experimental results are then stored in public/private databases where every lab can access and analyze multi-center and multi-omics data profiles. The data are integrated to predict clinical phenotype (subtype, prognosis, diagnosis) and help doctors for future patient stratification through specific biomarkers. Finally, this will lead to high accuracy for drug selection and personalized treatment.

References

1. Rudnick, P. A. *et al.* A Description of the Clinical Proteomic Tumor Analysis Consortium (CPTAC) Common Data Analysis Pipeline. *J. Proteome Res.* **15**, 1023–1032 (2016).
2. The Cancer Genome Atlas Research Network *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **45**, 1113–1120 (2013).
3. Cerami, E. *et al.* The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data: Figure 1. *Cancer Discov.* **2**, 401–404 (2012).
4. Vasaikar, S. V., Straub, P., Wang, J. & Zhang, B. LinkedOmics: analyzing multi-omics data within and across 32 cancer types. *Nucleic Acids Res.* **46**, D956–D963 (2018).
5. Dienstmann, R. *et al.* Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. *Nat. Rev. Cancer* **17**, 79–92 (2017).
6. Voest, E. E. & Bernards, R. DNA-Guided Precision Medicine for Cancer: A Case of Irrational Exuberance? *Cancer Discov.* **6**, 130–132 (2016).
7. Du, W. & Elemento, O. Cancer systems biology: embracing complexity to develop better anticancer therapeutic strategies. *Oncogene* **34**, 3215–3225 (2015).
8. Sever, R. & Brugge, J. S. Signal transduction in cancer. *Cold Spring Harb. Perspect. Med.* **5**, (2015).
9. Seliger, C. *et al.* Lactate-modulated induction of THBS-1 activates transforming growth factor (TGF)-beta2 and migration of glioma cells in vitro. *PloS One* **8**, e78935–e78935 (2013).
10. Røsland, G. V. *et al.* Epithelial to mesenchymal transition (EMT) is associated with attenuation of succinate dehydrogenase (SDH) in breast cancer through reduced expression of SDHC. *Cancer Metab.* **7**, (2019).
11. Moser, C. *et al.* STAT5b as molecular target in pancreatic cancer--inhibition of tumor growth, angiogenesis, and metastases. *Neoplasia N. Y. N* **14**, 915–925 (2012).
12. Pérez-Moreno, M. A. *et al.* A New Role for E12/E47 in the Repression of *E-cadherin* Expression and Epithelial-Mesenchymal Transitions. *J. Biol. Chem.* **276**, 27424–27431 (2001).
13. Blum, W. & Schwaller, B. Calretinin is essential for mesothelioma cell growth/survival in vitro: A potential new target for malignant mesothelioma therapy?: Putative role of calretinin in mesothelioma cell growth. *Int. J. Cancer* **133**, 2077–2088 (2013).
14. Markosyan, N. *et al.* Tumor cell–intrinsic EPHA2 suppresses antitumor immunity by regulating PTGS2 (COX-2). *J. Clin. Invest.* **129**, 3594–3609 (2019).
15. Collisson, E. A. & Olive, K. P. Pancreatic Cancer: Progress and Challenges in a Rapidly Moving Field. *Cancer Res.* **77**, 1060–1062 (2017).
16. Garrido-Laguna, I. & Hidalgo, M. Pancreatic cancer: from state-of-the-art treatments to promising novel therapies. *Nat. Rev. Clin. Oncol.* **12**, 319–334 (2015).
17. Neoptolemos, J. P. *et al.* Therapeutic developments in pancreatic cancer: current and future perspectives. *Nat. Rev. Gastroenterol. Hepatol.* **15**, 333–348 (2018).
18. Cohen, J. D. *et al.* Combined circulating tumor DNA and protein biomarker-based liquid biopsy for the earlier detection of pancreatic cancers. *Proc. Natl. Acad. Sci.* **114**, 10202–10207 (2017).
19. Cohen, J. D. *et al.* Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* **359**, 926 (2018).
20. Brancaccio, M., Natale, F., Falco, G. & Angrisano, T. Cell-Free DNA Methylation: The New Frontiers of Pancreatic Cancer Biomarkers' Discovery. *Genes* **11**, 14 (2019).

21. Liggett, T. *et al.* Differential methylation of cell-free circulating DNA among patients with pancreatic cancer versus chronic pancreatitis. *Cancer* **116**, 1674–1680 (2010).
22. Liu, X.-D. *et al.* Early detection of pancreatic ductal adenocarcinoma using methylation signatures in circulating tumour DNA. *Ann. Oncol.* **30**, v261–v262 (2019).
23. Best, M. G. *et al.* RNA-Seq of Tumor-Educated Platelets Enables Blood-Based Pan-Cancer, Multiclass, and Molecular Pathway Cancer Diagnostics. *Cancer Cell* **28**, 666–676 (2015).
24. Best, M. G., In 't Veld, S. G. J. G., Sol, N. & Wurdinger, T. RNA sequencing and swarm intelligence–enhanced classification algorithm development for blood-based disease diagnostics using spliced blood platelet RNA. *Nat. Protoc.* **14**, 1206–1234 (2019).
25. Best, M. G. *et al.* Swarm Intelligence-Enhanced Detection of Non-Small-Cell Lung Cancer Using Tumor-Educated Platelets. *Cancer Cell* **32**, 238-252.e9 (2017).
26. Shaikh, A. R. *et al.* Collaborative Biomedicine in the Age of Big Data: The Case of Cancer. *J. Med. Internet Res.* **16**, e101 (2014).
27. Zhang, C. *et al.* Systematically linking tranSMART, Galaxy and EGA for reusing human translational research data. *F1000Research* **6**, 1488 (2017).
28. Cancer Research Data Commons Repositories | CBIIT. https://datascience.cancer.gov/data-commons/repositories.
29. The European Bioinformatics Institute < EMBL-EBI. https://www.ebi.ac.uk/.
30. ELIXIR | A distributed infrastructure for life-science information. https://elixir-europe.org/.
31. Tsai, C. J., Riaz, N. & Gomez, S. L. Big Data in Cancer Research: Real-World Resources for Precision Oncology to Improve Cancer Care Delivery. *Semin. Radiat. Oncol.* **29**, 306–310 (2019).

Chapter 13

ENGLISH SUMMARY
NEDERLANDSE SAMENVATTING
RIASSUNTO IN ITALIANO

ENGLISH SUMMARY

The most prevalent type of pancreatic cancer is pancreatic ductal adenocarcinoma (PDAC). Morphologically, PDAC presents as a solid poorly-defined mass, typically between 2–4 cm at diagnosis, which infiltrates surrounding structures (peripancreatic adipose tissue, duodenum, stomach, portal vein, and regional lymph nodes). Moreover, PDAC epithelial cancer cells are embedded in a prominent desmoplastic reaction, known as the "stroma" which is mainly constituted by cancer-associated fibroblasts (CAFs), immune cells (T & B cells, Natural Killer cells, tumor associated macrophages), blood vessels, extracellular matrix and a liquid milieu of cytokines, growth factors and exosomes. An orchestrated crosstalk between epithelial and stromal cells can stimulates proliferation and metastatic growth alongside drug resistance and tumor relapse.

At present, diagnosis and treatment at early and late stage, is challenging. Patients rarely exhibit early symptoms and most of these are non-specific (jaundice, sudden weight loss, fatigue). Moreover, this type of tumor do not have sensitive and specific markers to aid detection. Thus, more sensitive and specific diagnostic markers for PDAC are urgently needed, especially non-invasive biomarkers that can be detectable in accessible biofluids such as blood, urine and saliva.

Early diagnosis of PDAC is a crucial step to improve the efficacy of treatment and patient outcomes. Current guidelines mention to adopt systemic chemotherapy with multi-drugs regimens: 5-fluorouracil/leucovorin combined with irinotecan and oxaliplatin (FOLFIRINOX) and the combination of gemcitabine and nab-paclitaxel. However, these protocols are associated with high toxicity and modest response rates.

Thus, major efforts should be taken to identify both new treatments and new biomarkers to select subgroups of patients who may benefit from specific therapies. To this end, the analysis of multiple molecular profiles has emerged as an unbiased strategy for the identification of novel drug targets and predictive biomarkers.

This thesis aimed to describe the underlying mechanisms of PDAC analyzing different types of molecules at single or multiple -omics layers. The overall goal was to provide improved insights into PDAC biology and novel candidate markers for diagnosis, prognosis, and treatments.

PART ONE – DISCOVERY OF NOVEL BIOMARKERS FROM PROTEOMICS

The first part of this thesis described several protein markers to accurately stratify PDAC patients with good and poor prognosis after surgical resection. Those markers were analyzed from stromal and epithelial micro-dissected tissues, and bulk tissues, and successfully validated on tissue microarrays (TMA) by immunohistochemistry and immunofluorescence. Additionally, this part provided an extensive description of the state of the art of integrative analysis using phosphoproteomics.

In **Chapter 2** we reviewed state-of-the-art methods for analysis of kinase activities, the latest large-scale proteogenomics studies with phosphoproteomics data, and bioinformatics methods where phosphoproteomics data can be treated as another -omics layer. We have highlighted the importance of multi-omics data visualization and data portals to enable the incorporation of expert knowledge in data analysis.

In **Chapter 3** and **4** we explored the prognostic value of PDAC protein analyzing bulk tumor and microdissected tumor (LCM).

In **Chapter 3** was described the use of co-expression networks on bulk tumor tissue to provide insights in PDAC biology and discover new protein markers for PDAC prognosis after surgical resection. In this study we found that one module enriched for epithelial-to-mesenchymal transition (EMT) and metabolism was significantly associated with overall survival (*p*-value = 0.01). The prognostic value of three proteins (SPTBN1, KHSRP and PYGL) belonging to this module was confirmed by immunohistochemistry.

In **Chapter 4,** laser-capture microdissection (LCM) was coupled to LC-MS/MS to enrich for stromal and epithelial compartments in PDAC samples of clinically resectable patients. These analyses revealed tumor CALB2 and stromal COL11A1 as prognostic marker specific for each compartment. Moreover, proteomics profile of stromal compartment was recapitulating previous stromal subtypes evaluated by transcriptomics data.

PART TWO - INNOVATIVE ANALYSES OF MiRNAs AND RNAs

The second part of this thesis examined the important role of miRNAs and their modified forms as well as intron retained RNA transcripts to decipher the complex biology behind PDAC and exploit new potential miRNAs targets for PDAC diagnosis and treatment.

Chapter 5 provided an overview of targetable miRNAs involved in aberrant PDAC metabolism. Innovative approaches to tackle cancer metabolic aberrations, such as glycolytic inhibitors targeting lactate dehydrogenase-A (LDH-A), showed promising effects in preclinical models. Therefore, development of new drugs targeting LDH-A and biomarkers for these treatments (e.g. miRNAs), should shorten the gap between preclinical studies and personalized therapies.

Chapter 6 described the use of blood platelets to accurate distinguish PDAC patients from benign hepatobiliary lesions (AUC = 0.94). The classifier was constructed using the support vector machine (SVM) method, learning from spliced profiles of RNA of tumor educated platelets (TEPs). This method can be helpful for clinicians to understand the correct strategy to adopt at the time of diagnosis.

In **Chapter 7**, we explored the biology of TEPs using multiple -omics layers (smallRNAs, transcriptomics and proteomics). This study confirmed the splicing activity on PDAC platelets.

In particular, the differential regulation of SPARC in PDAC and benign platelets, make platelets an interesting source for diagnostic tools, together with the different miRNAs and proteomic profiles. Further understanding of platelets content could therefore provide a dynamic and powerful approach for specific diagnosis of PDAC, paving the way for early detection intervention strategies.

PART THREE – COMPUTATIONAL TOOLS IN PHARMACOLOGICAL STUDIES

The majority of PDAC patients are diagnosed at a late stage or when the tumor has already distant metastases. These patients are treated upfront with chemotherapy, but most patients do not respond to the treatment and development of acquired resistance is common. This emphasizes the importance of predictive biomarkers for treatment response and the need of new computational methods to define better drug combination strategies.

In **Chapter 8** we described actual strategies for drug combination using available software packages such as CalcuSyn and Combenefit and the need of machine learning algorithms to refine the choice of drug candidates.

In **Chapter 9** we evaluated the use of a transmembrane protein hENT-1 as a predictive biomarker for gemcitabine uptake and the role of molecular factors that can influence hENT-1. Improved knowledge on hENT-1 and these factors should help the identification of subgroups of patients who may benefit from specific therapies and overcome the limitations of traditional biomarker studies.

In **Chapter 10,** we reviewed several splicing factors aberrantly expressed in PDAC and their relationships with their target miRNAs. Targeting aberrant splicing and the reciprocal interaction with deregulated miRNA could therefore provide more effective therapeutic approaches to combat the complex biology of PDAC and its chemoresistant features.

In **Chapter 11** we assessed the efficacy of two inhibitors (Pladeniolide-B and E7107) of the splicing factor SF3B1 in PDAC cells. Both compounds significantly reduced cellular migration, and E7107 showed promising results also in vivo. Additionally, SF3B1 expression was significantly associated to overall survival and progression-free survival in the TMAs. These results support further studies in this novel therapeutic approach for PDAC treatment.

Chapter 12 concludes with a general discussion and prompt towards more collaboration to gain data power and more investigation on liquid biopsy to boost the use in the clinic.

NEDERLANDSE SAMENVATTING

De meest voorkomende type van alvleesklierkanker ontstaat in de afvoerbuisjes en wordt ook wel ductaal pancreasadenocarcinoom genoemd (PDAC). Morfologisch gezien bestaat PDAC uit een vaste massa die bij diagnose gemiddeld 2-4cm groot is, daarnaast kan de tumor uitzaaien naar omliggende organen en weefsels (met name peripancreatisch vetweefsel, de twaalfvingerige darm, de maag, poortader en regionale lymfeklieren). Bovendien zijn epitheliale PDAC kankercellen omringd door een desmoplastische matrix, ook wel bekend als 'stroma'. Stroma is gemaakt van een extracellulaire matrix en bestaat voornamelijk uit tumor-geassocieerde fibroblasten (CAF's), imuuncellen (zoals T- en B-cellen, 'Natural Killer'-cellen, cytokinen, tumor-geassocieerde macrofagen) en bloedvaten. Daarnaast bestaat het uit een vloeibare component, die cytokinen, groeifactoren en exosomen bevat. De interactie tussen epitheel- en stromacellen kan een grote rol spelen bij de proliferatie en uitzaaiing van PDAC. Daarnaast kan deze interactie ook een bijdrage leveren aan resistentie tegen medicijnen en recidief van de tumor.

Tot op heden is het nog steeds een uitdaging om de diagnose en behandeling te bepalen voor PDAC, zowel in een vroeg maar ook in een later stadium. Klachten ontstaan meestal pas als de tumor is uitgezaaid en zijn vaak niet-specifiek (zoals geelzucht, plotseling gewichtsverlies, vermoeidheid). Bovendien heeft dit type tumor geen specifieke biomarker om de opsporing te vergemakkelijken. Daarom is er een dringende behoefte om meer sensitieve en specifieke diagnostische biomarkers voor PDAC te vinden. Daarbij is ook te denken aan niet-invasieve methoden van vloeibare biopsie van bijvoorbeeld bloed, urine en speeksel. Een vroege diagnose van PDAC stellen is daarnaast ook cruciaal om de effectiviteit van de behandeling en de genezing voor de patiënt te verbeteren. Standaard therapie voor PDAC bestaat uit een combinatie van verschillende therapieën: 5-fluorouracil/leucovorine gecombineerd met irinotecan en oxaliplatine (FOLFIRINOX) of de combinatie van gemcitabine en nab-paclitaxel. Helaas geeft deze chemotherpaie vaak ernstige bijwerkingen en heeft het maar een minimale effectiviteit.

Het is dus van belang om nieuwe biomarkers te identificeren om subgroepen van patiënten te selecteren die wel baat kunnen hebben bij specifieke therapieën. Door middel van het analyseren van verschillende moleculaire tumorprofielen kunnen er mogelijk nieuwe therapieën en voorspellende biomarkers aan het licht komen.

Dit proefschrift is gericht op het beschrijven van de onderliggende mechanismen van PDAC door de analyse van verschillende moleculen uit een varieteit aan omics data. Het doel is om een beter inzicht te krijgen in de biologie van PDAC en nieuwe biomarkers te selecteren die van waarde kunnen zijn bij de diagnose, prognose en therapie.

DEEL 1 – IDENTIFICATIE VAN NIEUWE BIOMARKERS UIT PROTEOMICA

Het eerste deel van dit proefschrift beschrijft de stratificatie voor een goede of slechte prognose van PDAC door de identificatie van eiwitmarkers. Deze markers zijn afkomstig uit epitheliale en stromale microdissctie weefsel of bulkweefsel en succesvol gevalideerd op weefsel microarray (TMA) door middel van immunohistochemie en immunofluorescentie. Bovendien geeft dit deel een uitgebreide beschrijving van de techniek achter integratieve analyse op eiwitactiviteit (fosfo-proteomica).

In **Hoofdstuk 2** worden de nieuwste methoden voor analyse van kinase-activiteiten, de laatste grootschalige functionele proteomica studies met fofo-proteomica data en bioinformatica-methoden waarbij fofo-proteomica data behandeld kunnen worden als een andere omics-laag besproken. Het belang van het visualiseren van verschillende omics data en dataportals worden hierin benadrukt om uiteindelijk de integratie van vakkennis in data-analyse mogelijk te maken.

In **Hoofdstuk 3 en 4** hebben we de prognostische waarde van PDAC-eiwitten onderzocht door cellen uit de tumor bulk en laser microdissectie (LCM) te analyseren.

In **Hoofdstuk 3** wordt het gebruik van co-expressie netwerken op tumor bulk weefsel beschreven om inzicht te verschaffen over de biologie achter PDAC en om nieuwe eiwitmarkers te identificeren voor de prognose na chirurgische ingreep. In deze studie ontdekten we dat een verrijking in het proces epitheliale en mesenchymale transitie (EMT) en metabolisme significant geassocieerd is met de algemene overlevingskans(p-waarde = 0.01). De prognostische waarde van drie eiwitten (SPTBN1, KHSRP en PYGL)die betrokken zijn bij dit proces zijn bevestigd door immunohistochemische kleuring.

In **Hoofdtuk 4** is de techniek LCM gekoppeld aan vloeistofchromatografie – massaspectrometrie/massaspectrometrie(LC-MS/MS) om de stromale en epitheliale weefsels te verrijken afkomstig uit reseceerbare tumor in PDAC-samples. Uit deze analyse komt naar voren dat CALB2 in tumoren en stromale COL11A1 als prognostische marker kunnen dienen voor zowel stromale als epitheliale weefsel. Bovendien is het proteomica-profiel van het stromale weefsel bevestigt met bekende stromale subtypen die werden geëvalueerd door transcriptomica data.

DEEL 2 – INNOVATIEVE ANALYSES VAN microRNA EN RNA

In het tweede deel van dit proefzicht is er onderzoek gedaan naar de rol van microRNA's en hun gemodificeerde vormen evenals RNA-transcripten uit intronen om de complexe biologie achter PDAC te ontcijferen. Hiermee hebben we het belang van nieuwe potentiële microRNA targets naar boven gehaald die voor een betere diagnose en behandeling in PDAC kunnen zorgen.

Hoofdstuk 5 geeft een overzicht van microRNA's die als doelwit kunnen dienen voor therapie en betrokken zijn bij een afwijkend metabolisme in PDAC tumoren. Preklinische modellen laten alleen veel belovend effect zien bij innovatieve benaderingen om metabole afwijkingen van kanker aan te pakkken. Bijvoorbeeld door toediening van glycolytische remmers die zich richten op het melkzuurdehydrogenase-A(LDH-A). Hierdoor zou de ontwikkeling van nieuwe geneesmiddelen die gericht zijn op LDH-A en biomarkers voor deze behandelingen (bijv. microRNA) de kloof tussen preklinische studies en gepersonaliseerde therapieën moeten verkleinen.

Hoofdstuk 6 beschrijft dat het gebruik van bloedplaatjes PDAC-patiënten een onderscheid kan maken tussen PDAC en goedaardige hepatibilaire laesis (AUC 0,94). Het predictie model gebruikt de methode Support Vector Machine (SVM), waarbij het model leert van geknipte ('spliced') RNA afkomstig uit tumor-opgeleide bloedplaatjes (TEP's). Deze methode kan voor clinici nuttig zijn bij het kiezen van de juiste behandeling op het moment van diagnose.

In **Hoofdstuk 7** hebben we de biologie van TEP's onderzocht uit meerdere bronnen van omics data (smallRNA, transcriptomie en proteomica). Deze studie bevestigt de knip ('splicing') activiteit op PDAC-bloedplaatjes. In het bijzonder het verschil in differentiële regulatie tussen het gen SPARC in PDAC en

goedaardige bloedplaatjes, maakt bloedplaatjes een interessante bron voor diagnostische onderzoek in combinatie met microRNA en proteomica-profielen . Een beter begrip van bloedplaatjes zou daarom een krachtige benadering kunnen zijn voor een specifieke diagnose en vroege detectie van PDAC.

DEEL DRIE - COMPUTATIONELE METHODENIN FARMACOLOGISCHE STUDIES

De meeste PDAC-patiënten worden in een laat stadium gediagnosticeerd of wanneer er al uitzaaiing naar andere organen heeft opgetreden. Deze patiënten worden vooraf behandeld met chemotherapie, de meeste patiënten reageren niet op de behandeling en daarnaast treedt er vaak resistentie tegen chemotherapie op. Hierdoor is het van belang om biomarkers te vinden die kunnen voorspellen of een patiënt wel of niet zal reageren op de therapie. Daarnaast ook de voorspelling of de patiënt profijt kan hebben van bijvoorbeeld een combinatie therapie. Daarvoor is het van belang om nieuwe computationele methoden te ontwikkelen om betere combinatie-strategieën voor medicijnen te definiëren.

In **Hoofdstuk 8** zijn bestaande software programma's,zoals CalcuSyn en Combenefit, voor de bepaling van combinatie therapie beschreven. Daarnaast is de behoefte naar nieuwe machine-learning algoritmen om de keuze naar mogelijke nieuwe combinaties te verfijnen hierin ook beschreven.

In **Hoofdstuk 9** is onderzocht of het voorspellende biomarker hENT-1 een rol speelt in de opname van gemcitabine en of het de rol van moleculaire factoren kan beïnvloeden. Meer kennis over hENT-1 en deze factoren zou de identificatie van subgroepen in PDAC kunnen helpen en baat leveren voor specifieke therapieën. Hiermee overwinnen we ook de beperkingen van bestaande studies naar biomarkers.

In **Hoofdstuk 10** hebben we de verschillende "splicing" factoren die in PDAC tot expressie komen en de relatie tot de microRNA targets beschreven. De afwijkende RNA "splicing" en de wederzijdse interactie met deregulatie van microRNA kan leiden tot een effectievere behandeling om de complexe biologie achter PDAC te bestrijden. Daarnaast kan het ook werken tegen resistentie in chemotherapie.

In **Hoofdstuk 11** hebben we de werkzaamheid van twee remmers (pladeniolide-B en E7107) van de "splicing" factor SF3B1 in PDAC-cellen onderzocht. Zowel pladeniolode-B als E7107 toonde een significante vermindering van cellulaire migratie, daarnaast vertoont E7107 ook *in vivo* veelbelovende resultaten. Bovendien is de SF3B1-expressie significant geassocieerd met een verbetering van algemene overlevingskans en progressievrije overleving in TMA's. Deze resultaten zijn veelbelovend voor verdere studies in de therapeutische benadering voor potentiële nieuwe behandeling voor PDAC patiënten.

Hoofdstuk 12 wordt afgesloten met een algemene discussie en raden we aan tot meer samenwerking om meer data te kunnen verzamelen en daarmee het onderzoek goed te valideren. Dit zou uiteindelijk de kliniek kunnen stimuleren om voor de diagnose van PDAC meer gebruik te maken van vloeibare biopsie.

RIASSUNTO IN ITALIANO

L'adenocarcinoma duttale pancreatico (PDAC) è il tipo di cancro al pancreas più diffuso. Al momento della diagnosi, il PDAC si presenta come una massa solida e poco definita, tipicamente compresa tra 2-4 cm, che si infiltra nelle strutture circostanti (tessuto adiposo peripancreatico, duodeno, stomaco, vena porta e linfonodi regionali). Le cellule tumorali epiteliali di PDAC sono incorporate in una prominente reazione desmoplastica, nota come "stroma", costituita principalmente da fibroblasti associati al cancro (CAF), cellule immunitarie (cellule T e B, cellule Natural Killer, macrofagi associati al tumore), vasi sanguigni, matrice extracellulare e un ambiente fluido in cui sono presenti citochine, fattori di crescita ed esosomi. Particolari interazioni tra cellule epiteliali e stromali possono stimolare la proliferazione cellulare e la crescita metastatica, nonché sviluppare resistenza ai farmaci o causare una recidiva del tumore.

Attualmente, la diagnosi di questo tipo di tumore e il trattamento sia nella fase iniziale che in quella avanzata, è molto complessa. I pazienti raramente mostrano sintomi precoci e la maggior parte di questi sintomi non sono specifici (ittero, improvvisa perdita di peso, affaticamento). Inoltre, questo tipo di tumore manca di marcatori sensibili e specifici, che aiuterebbero la diagnosi precoce. Pertanto, urge la necessità di trovare marcatori diagnostici più sensibili e specifici per PDAC, in particolare biomarcatori non invasivi che possono essere rilevati in biofluidi accessibili come sangue, urina e saliva.
La diagnosi precoce del PDAC è un passaggio cruciale per migliorare l'efficacia del trattamento e l'esito del paziente. Le attuali linee guida suggeriscono l'adozione della chemioterapia sistemica con regimi multi-farmaci: 5-fluorouracile / leucovorin combinato con irinotecan e oxaliplatino (FOLFIRINOX) e la combinazione di gemcitabina e nab-paclitaxel. Tuttavia, questi protocolli sono associati ad alta tossicità e a tassi di risposta modesti.
Pertanto, la ricerca deve prestare maggiori sforzi nell'identificare sia nuovi trattamenti che nuovi biomarcatori per selezionare sottogruppi di pazienti che possono beneficiare di terapie specifiche. A tal fine, l'analisi di più profili molecolari (genomica, proteomica, metabolomica ecc…) è emersa come una strategia imparziale per l'identificazione di nuovi bersagli farmacologici e biomarcatori predittivi.
Questa tesi mira a descrivere i meccanismi biologici alla base del PDAC analizzando diversi tipi di profili molecolari sia singolarmente che a livello multiplo. L'obiettivo generale è fornire una migliore comprensione della biologia del PDAC e nuovi marcatori utili per la diagnosi, prognosi e trattamenti di questo tumore.

PARTE PRIMA - IDENTIFICAZIONE DI NUOVI BIOMARKERS DAI DATI DI PROTEOMICA
La prima parte di questa tesi descrive diversi markers proteici per stratificare accuratamente i pazienti PDAC con prognosi buona e sfavorevole, dopo resezione chirurgica. Tali marcatori sono stati analizzati da tessuti micro-dissezionati stromali ed epiteliali e da tessuti mixed (non

dissezionati). Successivamente, questi marcatori sono stati convalidati su microarrays di tessuti (TMA) mediante immunoistochimica e immunofluorescenza.
In aggiunta, questa parte ha fornito un'ampia descrizione dell'analisi integrativa utilizzando la fosfoproteomica.

Nel **Capitolo 2** abbiamo esaminato metodi all'avanguardia per l'analisi delle attività delle chinasi, discutendo degli ultimi studi di proteogenomica su larga scala con dati di fosfoproteomica e metodi bioinformatici di integrazione. Abbiamo evidenziato l'importanza della visualizzazione dei molteplici profili molecolari ed elencato i portali web attualmente in uso.

Nei **Capitoli 3 e 4** abbiamo identificato markers prognostici per PDAC utilizzando dati di proteomica per tumore non dissezionato (**Capitolo 3**) e tumore dissezionato (**Capitolo 4**).

Nel **Capitolo 3** è stato descritto l'uso di co-expression network sul tessuto tumorale non dissezionato, in modo da approfondire la biologia del PDAC e scoprire nuovi markers proteici per la prognosi dopo resezione chirurgica. In questo studio abbiamo scoperto che un modulo era costituito principalmente da proteine coinvolte nella transizione epiteliale-mesenchimale (EMT) e nel metabolismo, ed era significativamente associato alla sopravvivenza dei pazienti (valore p = 0,01). Il valore prognostico delle proteine SPTBN1, KHSRP e PYGL appartenenti a questo modulo è stato confermato dall'immunoistochimica.

Nel **Capitolo 4**, la microdissezione laser (LMD) è stata accoppiata alla spettrometria di massa per acquisire con maggiore qualità, le proteine specifiche dei compartimenti stromali ed epiteliali. L'analisi è stata condotta in campioni PDAC di pazienti clinicamente resecabili. Queste analisi hanno rivelato che le proteine CALB2 e il COL11A1 potrebbero essere utilizzate come markers prognostici specifici per il compartimento epiteliale e stromale, rispettivamente. Inoltre, il profilo proteomico del compartimento stromale ricapitola i sottotipi stromali precedenti identificati dai dati di trascrittomica.

PARTE SECONDA - ANALISI INNOVATIVE DI MiRNA E RNA
La seconda parte di questa tesi ha esaminato il ruolo importante dei miRNAs e delle loro forme modificate, nonché dei trascritti di RNA contenenti introni. Questa parte mirava a decifrare la complessa biologia alla base del PDAC e sfruttare nuovi potenziali bersagli dei miRNA per la diagnosi e il trattamento del PDAC.

Il **Capitolo 5** ha fornito una panoramica dei miRNA coinvolti nel metabolismo anomalo del PDAC che potrebbero essere inibiti da farmaci. Approcci innovativi per affrontare le anomalie metaboliche del cancro, come gli inibitori glicolitici che prendono di mira la lattato deidrogenasi-A (LDH-A), hanno mostrato effetti promettenti nei modelli preclinici. Pertanto, lo

sviluppo di nuovi farmaci mirati a LDH-A e biomarcatori per questi trattamenti (ad esempio miRNA), dovrebbe ridurre il divario tra studi preclinici e terapie personalizzate.

Il **Capitolo 6** ha dimostrato l'uso delle piastrine per distinguere accuratamente i pazienti PDAC dalle lesioni epatobiliari benigne (AUC = 0,94). Il classificatore è stato sviluppato tramite metodo SVM (Support Vector Machine), apprendendo dai profili di RNA contenenti introni delle piastrine tumorali (TEP). Questo classificatore potrebbe aiutare i medici nella decisione della strategia corretta da adottare al momento della diagnosi.

Nel **Capitolo 7**, abbiamo esplorato la biologia delle TEPs utilizzando più profili molecolari (smallRNA, trascrittomica e proteomica). Questo studio ha confermato l'attività di splicing sulle piastrine PDAC. In particolare, la regolazione differenziale di SPARC in PDAC e piastrine benigne, rendono le piastrine una fonte interessante per la diagnosi. Un approfondimento maggiore del contenuto delle piastrine potrebbe quindi fornire un approccio dinamico e robusto per la diagnosi specifica di PDAC, aprendo la strada a strategie di intervento per la diagnosi precoce.

PARTE TERZA - STRUMENTI COMPUTAZIONALI NEGLI STUDI FARMACOLOGICI

La maggior parte dei pazienti PDAC viene diagnosticata in una fase avanzata o quando il tumore è già metastatizzato. Questi pazienti vengono trattati con la chemioterapia, ma la maggior parte dei pazienti non risponde al trattamento e la probabilità di sviluppare resistenza al farmaco è purtroppo ancora molto alta. Ciò sottolinea l'importanza dei biomarcatori predittivi per la risposta al trattamento e la necessità di nuovi metodi computazionali per definire migliori strategie di combinazione di farmaci.

Nel **Capitolo 8** abbiamo descritto le strategie effettive per la combinazione di farmaci utilizzando i pacchetti software disponibili come CalcuSyn e Combenefit e la necessità di algoritmi di apprendimento automatico per affinare la scelta dei farmaci.

Nel **Capitolo 9** abbiamo valutato l'uso di una proteina transmembrana hENT-1, come biomarcatore predittivo per l'assorbimento di gemcitabina all'interno della cellula tumorale e il ruolo dei fattori molecolari che possono influenzare hENT-1. Una migliore conoscenza di hENT-1 e di questi fattori dovrebbe aiutare l'identificazione di sottogruppi di pazienti che possono beneficiare di terapie specifiche e superare i limiti degli studi sui biomarcatori tradizionali.

Nel **Capitolo 10**, abbiamo esaminato diversi fattori di splicing espressi in modo anomalo in PDAC e le loro relazioni con i loro miRNA target. Inibire miRNAs coinvolti nella de-regolazione di fattori di splicing potrebbe fornire approcci terapeutici più efficaci per combattere la complessa biologia del PDAC e le sue caratteristiche chemioresistenti.

Nel **Capitolo 11** abbiamo valutato l'efficacia di due inibitori del fattore di splicing SF3B1 (Pladeniolide-B e E7107) nelle cellule PDAC. Entrambi i composti hanno ridotto significativamente la migrazione cellulare ed E7107 ha mostrato risultati promettenti anche in vivo. Inoltre, l'espressione di SF3B1 era significativamente associata alla sopravvivenza globale e alla sopravvivenza libera da progressione (PFS) nei tissue microarrays (TMA). Questi risultati forniscono una base per ulteriori studi su questo nuovo approccio terapeutico per il trattamento di PDAC.

Il **Capitolo 12** si conclude con una discussione generale e la richiesta di una maggiore collaborazione tra i vari centri di ricerca in modo da svolgere ricerche mirate e su larga scala. Inoltre, maggiore attenzione della ricerca deve essere incentrata sulla biopsia liquida, che rappresenta il futuro per una diagnosi precoce e minimamente invasiva oltre che ad un veloce e quindi costante monitoraggio del paziente sotto terapia.

ADDENDUM

LIST OF PUBLICATIONS

ACKNOWLEDGEMENTS

CURRICULUM VITAE

LIST OF PUBLICATIONS

Computational Analysis of Phosphoproteomics Data in Multi-Omics Cancer Studies.
Mantini G, Pham TV, Piersma SR, Jimenez CR.
Proteomics. 2021 Feb;21(3-4):e1900312. doi: 10.1002/pmic.201900312. Epub 2020 Oct 8.

Omics Analysis of Educated Platelets in Cancer and Benign Disease of the Pancreas.
Mantini G, Meijer LL, Glogovitis I, In 't Veld SGJG, Paleckyte R, Capula M, Le Large TYS, Morelli L, Pham TV, Piersma SR, Frampton AE, Jimenez CR, Kazemier G, Koppers-Lalic D, Wurdinger T, Giovannetti E.
Cancers (Basel). 2020 Dec 29;13(1):66. doi: 10.3390/cancers13010066.

Co-expression analysis of pancreatic cancer proteome reveals biology and prognostic biomarkers.
Mantini G, Vallés AM, Le Large TYS, Capula M, Funel N, Pham TV, Piersma SR, Kazemier G, Bijlsma MF, Giovannetti E, Jimenez CR.
Cell Oncol (Dordr). 2020 Dec;43(6):1147-1159. doi: 10.1007/s13402-020-00548-y. Epub 2020 Aug 29.

Microdissected pancreatic cancer proteomes reveal tumor heterogeneity and therapeutic targets.
Le Large TY, **Mantini G**, Meijer LL, Pham TV, Funel N, van Grieken NC, Kok B, Knol J, van Laarhoven HW, Piersma SR, Jimenez CR, Kazemier G, Giovannetti E, Bijlsma MF.
JCI Insight. 2020 Aug 6;5(15):e138290. doi: 10.1172/jci.insight.138290.

"Open Sesame?": Biomarker Status of the Human Equilibrative Nucleoside Transporter-1 and Molecular Mechanisms Influencing its Expression and Activity in the Uptake and Cytotoxicity of Gemcitabine in Pancreatic Cancer.
Randazzo O, Papini F, **Mantini G**, Gregori A, Parrino B, Liu DSK, Cascioferro S, Carbone D, Peters GJ, Frampton AE, Garajova I, Giovannetti E.
Cancers (Basel). 2020 Oct 31;12(11):3206. doi: 10.3390/cancers12113206.

Design, synthesis and biological evaluation of second-generation benzoylpiperidine derivatives as reversible monoacylglycerol lipase (MAGL) inhibitors.
Granchi C, Bononi G, Ferrisi R, Gori E, **Mantini G**, Glasmacher S, Poli G, Palazzolo S, Caligiuri I, Rizzolio F, Canzonieri V, Perin T, Gertsch J, Sodi A, Giovannetti E, Macchia M, Minutolo F, Tuccinardi T, Chicca A.
Eur J Med Chem. 2021 Jan 1;209:112857. doi: 10.1016/j.ejmech.2020.112857. Epub 2020 Oct 7.

Impact of hypoxia on chemoresistance of mesothelioma mediated by the proton-coupled folate transporter, and preclinical activity of new anti-LDH-A compounds.
Li Petri G, El Hassouni B, Sciarrillo R, Funel N, **Mantini G**, Zeeuw van der Laan EA, Cascioferro S, Avan A, Zucali PA, Zaffaroni N, Lagerweij T, Parrino B, Smid K, Deraco M, Granchi C, Braczko A, Smolenski RT, Matherly LH, Jansen G, Assaraf YG, Diana P, Cloos J, Peters GJ, Minutolo F, Giovannetti E.
Br J Cancer. 2020 Aug;123(4):644-656. doi: 10.1038/s41416-020-0912-9. Epub 2020 Jun 4.

Imidazo[2,1-b] [1,3,4]thiadiazoles with antiproliferative activity against primary and gemcitabine-resistant pancreatic cancer cells.
Cascioferro S, Petri GL, Parrino B, Carbone D, Funel N, Bergonzini C, **Mantini G**, Dekker H, Geerke D, Peters GJ, Cirrincione G, Giovannetti E, Diana P.
Eur J Med Chem. 2020 Mar 1;189:112088. doi: 10.1016/j.ejmech.2020.112088. Epub 2020 Jan 25.

CX-5461 Inhibits Pancreatic Ductal Adenocarcinoma Cell Growth, Migration and Induces DNA Damage.
El Hassouni B, **Mantini G**, Immordino B, Peters GJ, Giovannetti E.
Molecules. 2019 Dec 4;24(24):4445. doi: 10.3390/molecules24244445.

New avenues in pancreatic cancer: exploiting microRNAs as predictive biomarkers and new approaches to target aberrant metabolism.
Capula M, **Mantini G**, Funel N, Giovannetti E.
Expert Rev Clin Pharmacol. 2019 Dec;12(12):1081-1090. doi: 10.1080/17512433.2019.1693256. Epub 2019 Nov 24.

Uridine Cytidine Kinase 2 as a Potential Biomarker for Treatment with RX-3117 in Pancreatic Cancer.
El Hassouni B, Infante J, **Mantini G**, Ricci C, Funel N, Giovannetti E, Peters GJ.
Anticancer Res. 2019 Jul;39(7):3609-3614. doi: 10.21873/anticanres.13508.

To Combine or Not Combine: Drug Interactions and Tools for Their Analysis. Reflections from the EORTC-PAMM Course on Preclinical and Early-phase Clinical Pharmacology.
El Hassouni B, **Mantini G**, Li Petri G, Capula M, Boyd L, Weinstein HNW, Vallés-Marti A, Kouwenhoven MCM, Giovannetti E, Westerman BA, Peters GJ; EORTC PAMM Group.
Anticancer Res. 2019 Jul;39(7):3303-3309. doi: 10.21873/anticanres.13472.

Role of c-MET Inhibitors in Overcoming Drug Resistance in Spheroid Models of Primary Human Pancreatic Cancer and Stellate Cells.

Firuzi O, Che PP, El Hassouni B, Buijs M, Coppola S, Löhr M, Funel N, Heuchel R, Carnevale I, Schmidt T, **Mantini G**, Avan A, Saso L, Peters GJ, Giovannetti E.

Cancers (Basel). 2019 May 8;11(5):638. doi: 10.3390/cancers11050638.

ACKNOWLEDGEMENTS

"None of us is as smart as all of us." - Ken Blanchard

This thesis is the result of the collaborative work of several people: from the professors who first believed in me to carry out this project, to the colleagues who helped me on this journey full of emotions. Undoubtedly, friends and family members who sustained me for the whole period cannot be missed.
Therefore, I want to express my sincere gratitude to all the people who have given me a boost to complete this thesis.

Professor Jimenéz, Beste **Connie**, thanks for believing in me first, and thanks for all your work- and life-related advice. With you was my first approach to the Dutch world of work. It was surprising to see how a healthy relationship between employer and employee can change a person's attitude and way of working. Thanks for your honest advice and for being just you.

Professor Wurdinger, Beste **Tom**, this journey started thanks to you when you first contacted me for a meeting. We talked about your innovative algorithm to detect cancer from blood platelets. Thank you so much for opening me the door to this world.

Dr. Giovannetti, Cara **Elisa**, è stato certamente un onore e un vero piacere lavorare con lei. Un'insegnante di vita a tutto tondo nonché un fiume in piena di idee e sempre pronta a motivarmi e spronarmi per andare avanti, cercando di risolvere tutti i problemi di percorso insieme. Disponibile in qualsiasi giorno ed a qualsiasi orario, non esistono giorni di festa quando c'è in ballo la ricerca. La tua grinta e pazienza sono stati sicuramente un modello fondamentale che mi ha spronato ad andare avanti e a raggiungere questo importante traguardo. Grazie.

Dr. Bijlsma, Beste **Maarten**, even from the "AMC-locatie", you have always been present. You were carefully checking all my work and replaying efficiently to all the emails! I appreciated working with you, bedankt!

There are some take-home messages from my supervisors that I'll bring with me forever: the positive vibe and networking of **Tom**, the constant seeking of the truth of **Connie**, the perseverance and team-building of **Elisa,** and the elasticity of **Maarten**.

A big thank goes to the promotion commission: *Prof. Dr. J. Heringa*, *Prof. Dr. Renske*, *Prof. Dr. F. Beltram*, *Dr. T.V. Pham*, *Dr. A. Frampton*, *Dr. C. Bachas,* for accepting to review my thesis.

Another big thank goes to **Thang**, that despite the millions of research projects he was involved in, he tried to be present in every step of my research, giving me the proper feedback at the right time.
Thanks to the OPL core team: Alex, Thang, Sander, and Jaco, because all my work would not have been done without you!
Thanks to my office mates from OPL and pharmacology lab ;)

Ayse ☺, I should thank God for having you next to my desk, especially during the first year of my PhD ;) I will never forget your sense of humor (especially when things were going wrong)!! :p, the hours of laughs when trying to find the best code to use for a simple but automated volcano plot, with our "fancy" variable names :p. (The day usually ended with a lot of beer to forget about the total waste of time!)

Frank. You are the one going for lunch at 11.50, and I was the one going at 13.30, you are the one "things need time" and I was the one "I don't have time", I am a believer and you are the "concrete" guy… anyway I have to say we managed very well to keep those things balanced between us, and between a beer and another ;)

Tim, unfortunately, I only had one year to enjoy your happiness and creativity (see dry-ice in coffee glass), but I was lucky enough to enjoy your BBQ at Franziska's!

Franziska, I will never forget your moving out party in Leiden! I really enjoyed making the spritz and I hope you are still using that recipe! :p

Andrea ☺, I will never forget the morning I found you dancing in the office. It was probably a sunny and warm day of spring :p. I'll miss the endless chat that always started discussing about work and always ended with …. A lot of chat!!
I'm missing already the voice behind me: "Giu…?". We shared a lot, from the research project to the group of friends… and I cannot be happier than this <3

Thanks to all OPL members!!

Of course, **Asli ☺,** thank you for being there and for listening to all my complaints (as only a good Italian can do :p). Thank you for always giving me the right push to go forward. You were just essential those years.

Many thanks to all members of the pharmacology lab: **Btissame**, **Ornella**, **Lenka**, **Alessandro**, **Pei Pei**, **Ornella**, **Giovanna,** and **Giulia Bononi**. I enjoyed every single minute spent with you, both in the lab and outside, playing laser game (Lenka, you almost destroyed my nose :p), eating sushi, cheeseburgers, and pizzas ;)

Thanks to **Tessa**, I think I've learned a lot from you, so thanks for your work- and life-related tips.

Thanks to **Laura** that made my days with her big smile! Never seen a PhD student always happy and cheerful (even on rainy days). Ciao bellasssss!!

Thanks to **Danijela**, for your proper supervision on the project, for the chat at the coffee corner and on skype (in covid times).

Thanks to **Silvia** and **Mafalda**, professionals at the CCA, and always up for partying!!

Thanks to **Sjors,** you are probably one of the first CCA students I met, thanks for all your help.

Thanks to **Mjriam,** the human database of TV series. Thanks for being my mate of conferences (in Verona and Stockholm), thanks for giving me a bed when I was in Pisa, and for feeding me and let me feel home all the time!

Thanks to **Paolo Aretini** for your precious inputs and analysis ☺

Thanks to **Maartje** and **Simone,** always available for the endless bureaucratic problems!

Thanks to **Daoud** for helping me every time with the Zeus server! I will empty my folder, I promise!!!

Thanks to **Lucas** and **Hugo**! You were just lovely. Thanks for all the great moments we had together! (Lucas, I forgive you for my winter clothes' story!!)

From a house to another one, I want to thank my dear landlords **Nelleke** and **Edwin.** Thanks for being patient with us and for all the energy you gave us. Finding a couple like you, it's unique nowadays, and we are very honored to have met you guys!! (I will soon contact you about my new house furniture! :p)

…and now it's time to thank the Italian family in Amsterdam.

Per prima cosa, vorrei ringraziare **Linda,** perché senza di lei questo gruppo non sarebbe mai nato!

La seconda persona, portabandiera del gruppo, è **Paolo,** inutile dire, ma senza di lui il gruppo non avrebbe la stessa angolatura :p

Grazie a **Natascia** che si è trasferita ad Almere per amore portandosi con sé le domeniche di carbonara! Ahhhhh mannaggia Natascia!!! Grazie per averci dato da mangiare e da bere, ma soprattutto grazie per la tua infinita positività e allegria! Obviously she couldn't choose a BBQ chef as boyfriend!! Thanks **Patrick** for the BBQ time together!!

Grazie a **Luigi** che ha allietato i pomeriggi di lockdown suonando la chitarra, per la sua infinita allegria e insaziabile voglia di distruggermi casa! :p Aspettando la band e giggiplotta!!!!

Grazie ad **Erika** che nonostante portava avanti un dottorato abbastanza intenso, ha sempre trovato il modo di mandarmi un messaggio e di vederci. Sei stata presente ad ogni nostro cambiamento :p Grazie per le serate ad Utrecht e per i super pranzi domenicali ;) Grazie anche al tuo mate, **Francisco** che mi ha sempre trasmesso tranquillità e con una frase detta tra un pranzo e l'altro mi ha insegnato a prendere di petto i problemi della vita e farne dei propri punti di forza. Thank you guys!

About two years ago, **Francesca** and **Jeff** landed in Amsterdam from Washington DC. Jeff, the American boy with his legendary accent. **Francesca**, all those words won't be enough to say "Thank you". Ci siamo imparate a conoscere (si può dire ?) un po' per necessità. In questi mesi di pandemia, le nostre camminate sono state la nostra salvezza senza cui non sarei riuscita a scrivere gli ultimi articoli e concludere la mia tesi. Sono state camminate di risate, di sfoghi, di racconti d'infanzia, di organizzazioni del fine settimana, della spesa del sabato! :p. Si, perchè la nostra unica attività, come molti penso in questo periodo di pandemia, è stata la spesa del weekend (Non dimenticherò mai le settimane di programmazione per il Thanksgiving day!) Del resto siamo due persone opposte, tu meticolosa nell'organizzazione e nel rispettare i tempi, io quella disorganizzata sempre in ritardo che dimentica tutto (ma tanto c'è Francesca :p). Grazie di cuore tesò:*

Da una Francesca all'altra, o meglio alle altre! Ringrazio tutte le mie Francesc*he* italiane:

F. Colelli, *Stellì*, nonostante tutti questi anni di distanza sei sempre riuscita a starmi vicina con una chiamata ed in persona. Questo è un anno importante per me quanto lo è per te, anzi per voi! Un grazie va anche a **Bruno** che ha affrontato ben 2 viaggi in aereo per venire fin qui (cosa si è costretti a fare per amore della propria ragazza -.-). Grazie per le giornate passate qui ad Amsterdam, a Pescara e a Campo di Giove!! Grazie ragazzi!

F. Bronzi, *Ciccia,* ho il cuore pieno di gioia pensando a te. Il primo anno di questo dottorato non è di certo stato un anno semplice, forse nemmeno il secondo, il terzo ed il quarto. Ma posso dire che sei stata per me un esempio di motivazione ad andare avanti, quindi, se sono qui ora a scrivere queste pagine è anche grazie a te cara, Grazie Ciccia!!!!!
Poi ovviamente come non ringraziare **Leonard** con il suo fantastico " A' Bronzi" che riecheggia ancora in questa casa! Dai che presto torniamo a festeggiare!!

F. Ferrante, *Sapo*, che dire… ci siamo divertite con Victoria al De School quella notte eh! :p Ne abbiamo fatte tante e ovunque! Mi hai accompagnata via FaceTime in questo viaggio, ricordandoci tutti i momenti passati insieme, per poi rivederci l'estate a Pescara. Ci siamo conosciute ragazzine ad un concerto e siamo ancora qui, donne che si credono in carriera :p

Grazie a **Silvia**, *Silvietta,* che ormai si è stabilizzata con **Francesco** diventando il mio punto di riferimento a Bologna. Ah Silviè! Alla fine sei stata la più veloce! Comunque non dimenticherò mai quel giorno che aspettavo te e Francesca a casa qui ad Amsterdam e siete arrivate dopo mezz'ora, ho aperto la porta ed eravate bagnate dalla testa ai piedi, con Francesca che aveva staccato un pezzo della bicicletta! Avete sicuramente assaporato la vera vita olandese ;)

Grazie anche a **Simona** e **Simone** che erano pronti a venire ad Amsterdam ma poi la pandemia li ha costretti a rimandare tutto. Grazie Simona perché sei una delle poche persone che cercano di dirti le cose come stanno e grazie per averci riportate a casa sane e salve il giorno di ferragosto da Scanno! :p Grazie Simone per le serate passate insieme e per i consigli sulla chitarra.

Un Grazie va anche **Cips** e **Sara**: Lorenzo, non finirò mai di ringraziarti per tutte le volte che mi hai "salvato il sedere" (l'ultima di recente in zona rossa). Spero tu non abbia salvato il mio numero con "fatti un favore: non rispondere". Grazie a **Sara** che ci ha intrattenuto in questi mesi di pandemia con le varie ricette di cac e' ove (però poldino non funziona sarè :p). Grazie!

Un grazie va anche a **Micaela**! Che mi ha s(u/o)pportata in questi mesi di pandemia tra un esercizio e una chiacchiera! Grazie Micky! (Però il jumping jack deve essere abolito dal riscaldamento!!!)

Grazie a tutti quelli che ho incrociato in questo viaggio, soprattutto se siete riusciti a strapparmi un sorriso! ;)

Grazie a zia **Antonella** e zio **Giovanni** perché mi hanno sempre supportata in questi anni di "estero", offrendomi un posto dove dormire nei miei scali a Bergamo. Grazie alla zia perché è stata ispirazione dei miei obiettivi di vita, e grazie allo zio perché si è sempre dimostrato interessato al mio lavoro.
Grazie alla mia cuginetta **Mariachiara**, il quale mi ha dimostrato che con la volontà tutto è possibile.

Grazie a **Chiara**, mia sorella, che prova a fare la sorella maggiore ma non ci riesce :p che ha provato a portarmi una bottiglia di vino dall'Italia ma è "accidentalmente" caduta in aeroporto a Eindhoven, che ha distrutto la mattinata di Lucas per farsi spiegare cosa volesse dire

“Indeed”, che ha scelto sempre i periodi migliori per venire ad Amsterdam (Novembre-Dicembre). Ma in fondo… chi se ne fregaaaaaa! E grazie!
Grazie alla super nonna! Nonna **Fausta,** notevole temeraria anche di fronte al covid-19! Le tue preghiere mi hanno sicuramente aiutata in questo cammino (oppure sono stati gli integratori? Chi lo sa!) Grazie per essere semplicemente te stessa ☺

Grazie a **mamma** e **papà**, perché ovviamente se sono arrivata fin qui è soprattutto grazie ai vostri sacrifici lavorativi e personali. Grazie per aver creduto in me e per avermi dato la libertà di scegliere. Non siete genitori comuni, non lo siete mai stati e mi ritengo fortunata per questo.

Infine ci sei tu, *ciccio*, il motore di tutto questo percorso. Un percorso difficile certo, che senza il tuo costante sostegno (e sottolineo costante), non si sarebbe concluso. Ci siamo gestiti bene è vero, soprattutto durante la pandemia, dove mi hai letteralmente viziata: la colazione pronta al mattino, la cucina del pranzo o della cena, i pacchi di biscotti comprati che non sei riuscito nemmeno ad aprire… e poi la domanda mattutina del fine settimana “Cosa vuoi fare oggi?”. Mi hai sempre appoggiata in qualsiasi idea, stupida o buona che era, anche se magari avresti preferito fare altro :p. Mi hai seguita in posti e fatto cose che forse non avresti mai voluto fare (vedi il kitesurf). Grazie per farmi sentire importante ogni giorno e per aver reso questa esperienza più leggera… sai benissimo che non ce l’avrei fatta senza di te! *Love u*!

L'ultimo saluto va alle persone che sono venute a mancare durante questo percorso ed In particolar modo ai miei nonni Tommaso e Vincenzo, Grazie!

Curriculum Vitae

Giulia Mantini was born on the 22nd of October 1991 in Pescara, Italy. She grew up in the city, spending her spare time mainly with friends and practicing several sports such as swimming, wind-surfing, volleyball, skiing, and snowboarding. Thanks to her family, she developed the love to travel around. In 2010, she moved to Teramo to start the bachelor's degree in Biotechnology at the University of Teramo. In the last year of her bachelor studies, she got close to the Bioinformatic world thanks to an open conference given at the University. In 2014, she started the master's degree in Bioinformatics in Bologna, and in 2016 she moved to Stockholm (Sweden) for the master's internship. The internship lasted six months, during which she analyzes data from a novel multiplexing barcode protocol for NGS technologies (and went skiing on the Swedish mountains with her colleagues). This work was carried out at the SciLifeLab under the supervision of Prof. Mats Nilsson in the Molecular Diagnostic Group. The same year, she got a master's degree in Bioinformatics and continued to work in the same group in Stockholm as a research assistant. She then moved to Amsterdam (NL) in 2017, where she started the Ph.D. in computational biology at the Cancer Center of Amsterdam. Her project involved four labs based in the Netherlands and Italy, aiming to find new biomarkers and drug targets for pancreatic cancer through different omics data. Results of the Ph.D. project are described in this thesis. At the moment, she's enjoying one of the best sports that the Netherlands has to offer, kitesurfing.